COMPRESSIVE SENSING FOR SPECTRAL DOMAIN OPTICAL COHERENCE TOMOGRAPHY

by

Daguang Xu

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

October, 2014

Abstract

Spectral domain optical coherence tomography (SD OCT) imaging with high axial resolution and a large imaging depth requires a large number of sampling points in the spectral domain. This requires a high-resolution spectrometer with a large linear array camera which leads to a large amount of k-space measurements and a long data acquisition time that makes the imaging susceptible to unavoidable motion artifact. Furthermore such devices can be expensive and require high-speed electronics.

In this dissertation, compressive sensing (CS) SD OCT that reconstructs the images using only a portion of the k-space measurements required by the classical Shannon/Nyquist rate was proposed and studied. Several advanced CS SD OCT algorithms have been developed and evaluated. First, modified non-uniform discrete Fourier transform (MNUDFT) matrix was proposed, which enables CS SD OCT using under-sampled non-linear wavenumber spectral data. Second, the noise reduction using Modified-CS was studied which shows that the averaged Modified-CS SD OCT results in better image quality in terms of SNR, local contrast and contrast to noise ratio (CNR), compared to the classical averaging method. Third, a novel three-

ABSTRACT

dimensional (3D) CS SD OCT sampling pattern and reconstruction procedure was proposed. The novel 3D approach enables efficient volumetric image reconstruction using the k-space measurements under-sampled in all three directions and reduces the amount of required measurements to less than 20% of that required by regular SD OCT.

CS SD OCT is commonly solved by an iterative algorithm that requires numerous matrix-vector computation, which is computationally complex and time-consuming if solved on CPU-based systems. However, such computation is ideal for parallel processing with graphics processing unit (GPU) which can significantly reduce its computation time. In this dissertation, real-time CS SD OCT was developed on a conventional desktop computer architecture having three GPUs. The GPU-accelerated CS non-uniform in k-space SD OCT and real-time CS SD OCT with dispersion compensation were also proposed and implemented using the same computer architecture.

Thesis Advisor: Prof. Dr. Jin U. Kang (Dissertation Reader)

Committee Members: Prof. Dr. Trac Duy Tran (Dissertation Reader)

Prof. Dr. Mark Foster

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Jin U. Kang for giving me the opportunity to finish my PhD study at Johns Hopkins. Your guidance, generous support and encouragement enables me to conduct my dissertation research. Thank you for your great mentorship during my graduate study. Your commitment to excellence has been and will always be an invaluable influence to me.

I am also extremely grateful to my dissertation committee members, Dr. Trac Duy Tran, Dr. Mark Foster. I have greatly benefited from their expertise. I also would like to thank all of my great teachers at Johns Hopkins University, especially Dr. Xingde Li, Dr. Randal Burns, Dr. Bruno Jedynak for being my GBO committee members. In addition, I would like to thank current and previous students and postdocs in our lab: Dr. Xuan Liu, Dr. Mingtao Zhao, Dr. Cheol Song, Dr. Yong Huang, Dr. Yi Yang, Mr. Jaepyeong Cha, Mr. Gyeong Woo Cheon. I treasure their input and friendship.

I am also very grateful to all the professors and students at CLSP. I enjoy my experience there during the first and second year of my PhD study. I want to thank Dr.

ACKNOWLEDGMENTS

Damianos Karakos and Dr. Sanjeev Khudanpur for supporting me and supervising me during this period.

I also want to thank my mentor Dr. Michal Sofka and team leader Dr. Shaohua Kevin Zhou at SCR, for hosting me as an intern from 2013 to 2014 and extend to me the opportunity to work as a full-time employee. I also want to thank everyone in the team: Dr. Nathan Lay, Dr. David Liu, Dr. Xuefei Guan, Dr. Glenn Meredith, Dr. JinHyeong Park, Dr. Hien Nguyen, Dr. Bernhard, Geiger, Dr. Carol Novak, Dr. Noha El-Zehiry, Dr. Jingdan Zhang, Mr. Wen Jiang, Miss Pinar Firat.

I want to express my deep thank to my wife, Yuezhe Tang, for her unconditional love and her sacrifice for me during my PhD studies. Words cannot express my love and gratitude to you. I will spend the rest of my life loving, cherishing and protecting you. Also, my son, Risuning Xu. You are my impetus to move forward. I also want to thank my other family members in China for their love and support.

Dedication

To my wife, son and parents!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Compressive sensing spectral domain optical coherence tomography

Optical coherence tomography (OCT) [2–4] is widely studied as medical diagnosis and imaging systems in many fundamental research and clinical applications. By measuring the backscattered light, OCT can provide cross-sectional and volumetric imaging of a wide range of biological tissues with micro-meter spatial resolution [5,6] at up to multi-megahertz A-line imaging rate [7].

Compared to the original time domain OCT (TD OCT) [2] which uses a low-coherence light source and employs a variable scanning reference arm to coherently gate backscattered light from various depth in a sample, Fourier domain OCT (FD

OCT) [8, 9] has shown superior detection sensitivity and imaging speed [10–16]. FD OCT enables the reconstruction of the depth-profile (A-scan) from the Fourier transform of the interferometric signal generated by mixing backscattered or backreflected light with a reference light reflected from a fixed position. FD OCT can be performed in two ways: spectral-domain OCT (SD OCT) [8–10] which uses a broad-band light source and an optical spectrometer having a line-scan array detector such as a CCD or CMOS camera, or swept-source OCT (SS OCT) [8, 13, 15, 17, 18] that employs a tunable laser that rapidly tunes narrow-band light over a broad bandwidth, and a point detector that collects the spectral measurement whose wavenumber changes over time.

Currently, FD OCT image generation algorithm requires a large number of sampling points in the spectral domain beyond the Nyquist rate which results in a high sampling rate for images requiring both large depth and high-axial resolution [13]. In SD OCT, this in turn requires the use of a high-resolution spectrometer with a large linear array camera, while a laser source with a high digitizer rate is necessary in SS OCT. Such CCD/CMOS cameras and laser sources are usually expensive and significantly increases the data acquisition time. In clinical applications, FD OCT with a high speed is often desired, since longer acquisition time makes the imaging susceptible to the unavoidable motion artifact. In addition, it is challenging to transfer and process the large amount of data acquired to achieve high speed and high resolution OCT imaging [19, 20].

In the past decade, compressive sensing (CS) [21, 22] has emerged as an effective technique for sampling a signal with smaller number of measurements than that required by the classical Shannon/Nyquist theory. This novel method enables significant large reduction of the required sampling size and thus reduces data acquisition time. For an accurate reconstruction, CS requires that the signal is sparse in some domain that is incoherent to the sampling domain. The sparsity here indicates the number of coefficients that represent the signal that is close to 0.

CS has already had notable successes when applied to other medical imaging modalities, such as magnetic resonance imaging (MRI) [23–25], computed tomography (CT) [26, 27], diffuse optical tomography [28, 29]. Applications of CS in OCT imaging have been reported by many research groups [30–37] and high quality imaging has been obtained with a significantly reduced amount of spectral measurements compared to the Nyquist rate requirement.

Previous works on CS SD OCT require linear wavenumber k-space undersampling because a uniform discrete Fourier transform (UDFT) matrix is used as the sensing matrix. This is inherited from the conventional FDOCT image generation algorithms. However, in most FDOCT systems, the spectra are linear in wavelength and non-linear in wavenumber. Two methods have been proposed to obtain undersampled linear wavenumber data from the non-linear wavenumber spectra in SDOCT: the k-space grid filling [30] and the k-linear random mask [31]. These two methods, however, have their own limitations: the previous one requires a complicated and

time-consuming spectral calibration with numerical interpolation; the latter one also demands an inflexible pre-calibration process that needs to be repeated even if a slight change on the sampling rate is desired. It also has an upper bound on the under-sampling rate because of the nature of the non-uniformity of the wavenumber in SD OCT. In this dissertation, we proposed a novel CS method on SD OCT [38]. By replacing the widely used uniform discrete Fourier transform (UDFT) matrix with a new sensing matrix which is a modification of the non-uniform discrete Fourier transform (NUDFT) matrix, it is shown that the under-sampled non-linear wavenumber spectral data can be used directly in the CS reconstruction.

Dispersion in FDOCT introduces a frequency-dependent phase to the Fourier components which degrades the axial resolution and reduces the sensitivity [3]. Dispersion compensation methods have been successfully implemented by both hardware and software approaches [1,39–41]. However, dispersion compensation has never been discussed in the context of CS reconstruction of FDOCT signal. In this dissertation, we propose two ways to incorporate dispersion compensation in CS reconstruction. The first one is implemented by adding the dispersion correcting terms to the modified NUDFT matrix [38]; the second one is implemented by multiplying the dispersion correcting terms to the under-sampled spectral data before CS reconstruction. Dispersion compensation then becomes a by-product of CS reconstruction, which significantly improves the image quality as well as the axial resolution.

Due to the interferometric nature of OCT, speckle noise reduction has long been

a focus of OCT research. Various methods for speckle noise reduction have been proposed based on hardware and software approaches to enhance the quality of OCT image [42–44]. One very successful and classical method is through averaging intensity of a series of B-scans that are successively recorded with slightly lateral offset [44]. Typically, average of 10 B-scans recorded with total 300um lateral offset is sufficient to provide a satisfying high-quality image. Ubiquitous system noise is another major noise source. Averaging method is also known to be efficient at reducing such system noise. One limitation of the classical averaging method is the data acquisition speed and large data storage requirement. As stated above, CS requires only a fraction of the total spectral measurements to accurately recover an image, which can significantly reduce the data amount and increase the data acquisition speed. Since CS enhances signal's sparsity, it has been shown to be good at reducing noise [23]. One outstanding benefit for CS comparing to commonly used denoising methods, such as filters, is that no additional denoising procedure is needed. The noise reduction is a by-product of the reconstruction procedure. In this dissertation, we explored the noise reduction in SD OCT using the Modified-CS [45, 46]-based averaging method and show that the averaged Modified-CS reconstructions achieves better image quality in terms of signal-to-noise-ratio (SNR), local contrast, and contrast-to-noise-ratio (CNR), compared to the calssical averaging method while reducing the total amount of spectral measurements required to reconstruct the image. The same is also true when compared to standard CS based averaging method with the same amount of

under-sampled data.

Ultrahigh resolution volumetric optical coherence tomography (OCT) imaging could become a powerful and efficient technique for a wide range of medical applications [47–50]. The volumetric image is usually obtained by assembling multiple consecutively acquired B-scans. Compared to the B-scan imaging, it has several advantages [7] such as volumetric visualization and quantification of tissue morphology features, reduction of the sampling error introduced by missing a sample location in 2D imaging, and reconstruction of en-face view of the tissue. Nevertheless, volumetric OCT imaging requires a long data acquisition time and large amount of spectral measurements which makes it susceptible to the unavoidable motion artifact and increases the discomfort of the patient in the clinical settings. Thus, we propose a novel sampling pattern that under-samples the spectral data for the volumetric OCT image in all three directions (axial direction, fast-scanning lateral direction and slow-scanning lateral direction), which results in less than 20% of the measurements required by the Shannon/Nyquist theory. The under-sampled data is introduced into a three-step CS strategy for recovering the volumetric image.

1.2 Real-time CS SD OCT based on GPU

CS facilitates the minimal data acquisition, increases imaging speed and decreases the storage and transfer bandwidth needs. However, regardless of which CS recon-

struction algorithm is used, CS SD OCT reconstruction takes significantly more time compared to the reconstruction of regular OCT image that can be obtained at more than 80K A-line speed [19, 51]. CS SD OCT is usually solved by an iterative algorithm that requires numerous matrix-vector computation, which is computationally complex and time-consuming if solved on the CPU based systems. Most of the current CS SD OCT studies are implemented on the "post-processing" mode, which obtains the under-sampled spectral data and saves them to the computer hardware; then a separate program reads in the under-sampled data, does the CS reconstruction, and displays the results. In addition, most of the current CS SD OCT studies are based on the Matlab programs that are much slower than the compiled programs. This has been the major hinderance to implement CS in the clinical settings that typically require real-time imaging acquisition, reconstruction and visualization simultaneously.

The most time-consuming step in CS is the matrix-vector multiplication. However, such computation is ideal for parallel computing, which can significantly reduce the computation time. Various methods based on parallel computing have been developed to accelerate the CS reconstruction, such as field-programmable gate array (FPGA) [52, 53], multi-core CPU parallel processing [54–56], and graphics processing unit (GPU) [53,55,57]. Compared to FPGA and multi-core CPU methods, GPU provides massively parallel execution resources and high memory bandwidth. In addition, GPU acceleration is more cost-effective and easier to be integrated into current OCT system. GPU has been widely used in processing regular SD OCT signals [19,41,58,59]

which achieved in excess of more than 320k Ascans/s with A-scan having 2048 pixels.

Prior to 2006, GPU has very limited use in the science and engineering applications, due to its limited support for both IEEE floating-point standards and software programming. On Feb 15 2007, NVIDIA release the initial software development kit (SDK) for the Compute Unified Device Architecture (CUDA), which gives the programmer direct access to the memory of the parallel computing elements of GPU. Since then, various applications in many fields have benefited from the acceleration of GPU, which often results in several order of performance improvement over the previous state-of-the-art implementations. The schematic structure of GPU is demonstrated in Fig. 1.1 [60]. The GPU contains several multiprocessors, which consists of multiple scalar processor (SP) cores, a multithread instruction unit and on-chip shared memory. The multiprocessor creates, manages, and executes concurrent threads which execute in batches of 32 threads called warps, providing single instruction, multiple data (SIMD) style parallelism. The NVIDIA GeForce GTX 670's specification is shown in Fig. 1.2. It is one of the GPUs used in our processing system for CS SD OCT, which has 1344 stream processors, 980 MHz processor clock and 2GB global memory.

In this dissertation, we implemented real-time CS SD OCT with the undersampled linear wavenumber sampling using a conventional desktop computer architecture having three GPUs to achieve a faster acceleration. The GPU-accelerated non-uniform fast Fourier transform (NUFFT)-based CS SD OCT and the GPU-

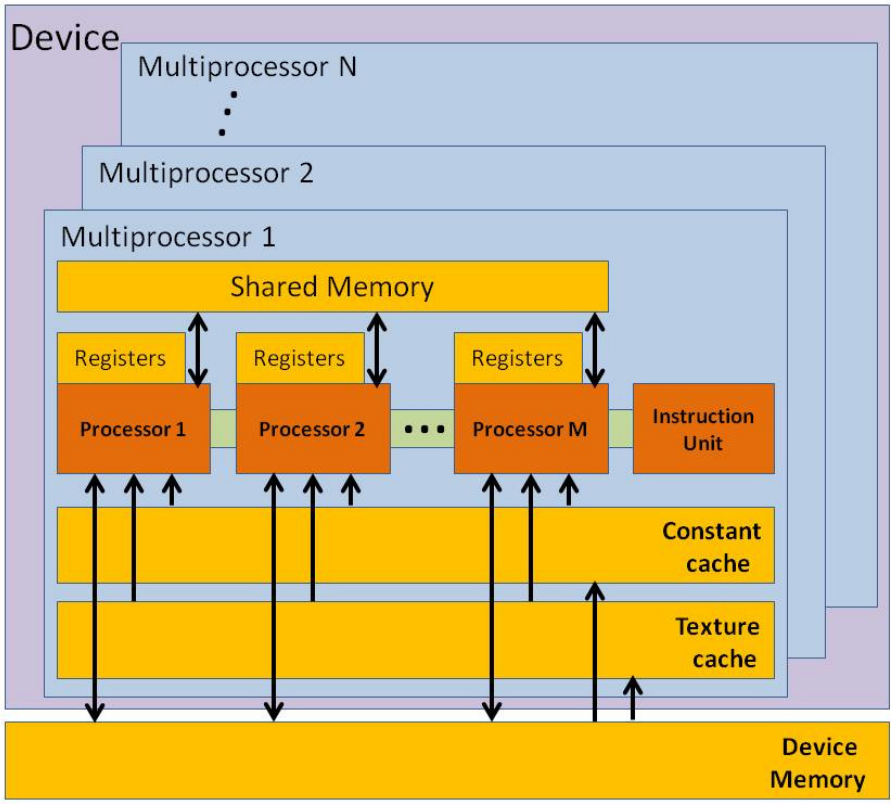

Figure 1.1: Schematic GPU architecture.

accelerated modified non-uniform discrete Fourier transform (MNUDFT)-based CS SD OCT using the under-sampled non-linear wavenumber sampling were also developed on the same architecture. Finally, we implement real-time CS SD OCT with dispersion compensation using the under-sampled linear wavenumber sampling.

1.3 Organization of Dissertation

The rest of this dissertation is organized as follows. Chapter 2 reviews the principle of OCT and its signal processing. Chapter 3 covers an overview of CS theory. Chapter

Nvidia Geforce GTX 670	
CUDA Cores	1344
Graphics Clock (MHz)	915
Processor Clock (MHz)	980
Memory (GB)	2
Memory Clock (Gbps)	6
Mmeory Bandwidth(GB/sec)	192.2
Maximum Graphics Card Power (W)	170
Supplementary Power Connections	Two 6-pin

Figure 1.2: Nvidia Geforce GTX 670 specification

4 describes the modified non-uniform discrete Fourier transform (MNUDFT)-based CS SD OCT with dispersion compensation. Chapter 5 covers a validation of GPU-accelerated real-time CS SD OCT using different transformation and under-sampled data, as well as implementation of real-time CS SD OCT with dispersion compensation. Chapter 6 demonstrates noise reduction using Modified-CS SD OCT and volumetric CS SD OCT imaging with a novel sampling pattern and reconstruction strategy. Finally, Chapter 7 summarizes the dissertation and discusses briefly the future directions of this work.

Chapter 2

Principles of Optical Coherence Tomography

In this section, basic principles of OCT are reviewed. First, the concept of optical coherence and interference which forms the foundation of OCT imaging is described. This is followed by a description of the basic principles of OCT systems, signal analysis and processing of FD OCT, as well as basic properties of FD OCT, such as resolution, imaging range and imaging sensitivity. Finally, the chapter concludes with a description of the optical dispersion and compensation in SD OCT.

2.1 Optical Coherence

Proved first by the Thomas Young's double-split experiment, optical coherence describes the property of light as optical waves that enables stationary interference. There are two different types of coherence in optics: spatial coherence and temporal coherence. The spatial coherence describes the correlation between two optical waves at two different positions in space, while the temporal coherence demonstrates the correlation of two optical waves at the same position but with a time difference. OCT is based on the temporal coherence of an optical wave and a time-shifted version of itself at a spatial position.

Denote an optical wave at spatial position $\mathbf{r}$ with time t as $A(\mathbf{r}, t)$, which is a complex function. The mutual coherence of two optical waves at position $\mathbf{r}_1$ and $\mathbf{r}_2$ with time t_1, t_2, respectively can be expressed as:

$$C(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2) = \langle A^*(\mathbf{r}_1, t_1) A(\mathbf{r}_2, t_2) \rangle \tag{2.1}$$

where $\langle \ldots \rangle$ is the averaging operation over the observation time. $A^*(\mathbf{r}_1, t_1)$ is the complex conjugate of $A(\mathbf{r}_1, t_1)$. The degree of coherence is usually used to quantify the coherence of two optical waves which is the normalization of the mutual coherence in Eq. (2.1):

$$c(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2) = \frac{\langle A^*(\mathbf{r}_1, t_1) A(\mathbf{r}_2, t_2) \rangle}{\sqrt{\langle A^*(\mathbf{r}_1, t_1) A(\mathbf{r}_1, t_1) \rangle} \sqrt{\langle A^*(\mathbf{r}_2, t_2) A(\mathbf{r}_2, t_2) \rangle}} \tag{2.2}$$

where

$$\begin{aligned} c(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2) &= 0 \quad \text{complete incoherent} \\ 0 < |c(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2)| &< 1 \quad \text{partial coherent} \\ |c(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2)| &= 1 \quad \text{complete coherent} \end{aligned} \tag{2.3}$$

For temporal coherence, $\mathbf{r}_1$ is the same as $\mathbf{r}_2$. Denote the time different as $\tau = t_2 - t_1$. The temporal coherence and its degree are:

$$C_t(\tau) = \langle A^*(t)A(t+\tau)\rangle \tag{2.4}$$

$$c_t(\tau) = \frac{\langle A^*(t)A(t+\tau)\rangle}{\langle A^*(t)A(t)\rangle} \tag{2.5}$$

The coherence time τ_c of a light source is approximately given by:

$$\tau_c = \frac{1}{\Delta\nu} \approx \frac{\lambda^2}{v_c \Delta\lambda} \tag{2.6}$$

where $\Delta\nu$ and $\Delta\lambda$ are the bandwidth of the light source in wavenumber and wavelength, respectively. λ is the central wavelength of the light source. v_c is the speed of light in vacuum.

The corresponding coherence length is defined as:

$$l_c = v_c \tau_c = \frac{\lambda^2}{\Delta\lambda} \tag{2.7}$$

A light source with a broad bandwidth, which implies a short coherence length and

coherence time, is usually used in the OCT system.

2.2 Interferometry and Interferometers

Since Young's establishment of the wave theory of light and extended by Fresnel, optical interferometry has long been used as an effective technique to extract useful information from the magnitude and phase of the interference pattern between two superimposed waves. To achieve stationary interference, these two light waves have to have almost the same frequency, the same polarization and a constant phase difference. In real applications, they are usually generated by splitting the light beam from a coherent light source.

Michelson interferometer is one of the most common configurations of optical interferometer. The light from a coherent light source is splitted into two light beams which travel different path to introduce a phase difference. A schematic configuration of Michelson interferometer is shown in Fig. 2.1. The light emitted from the light source hits the beam splitter and is divided into two different paths: one goes to mirror 1 while the other travels to mirror 2. The reflected lights from both paths are recombined after travelling through the beam splitter again to produce an interference pattern. Finally, the interference of the lights is measured using a detector. Denote the light travels to mirror 1 as $A_1(t)$ and the other one that travels to mirror 2 as $A_2(t + d/v_c)$. d is the optical path length (OPL) difference. The time difference

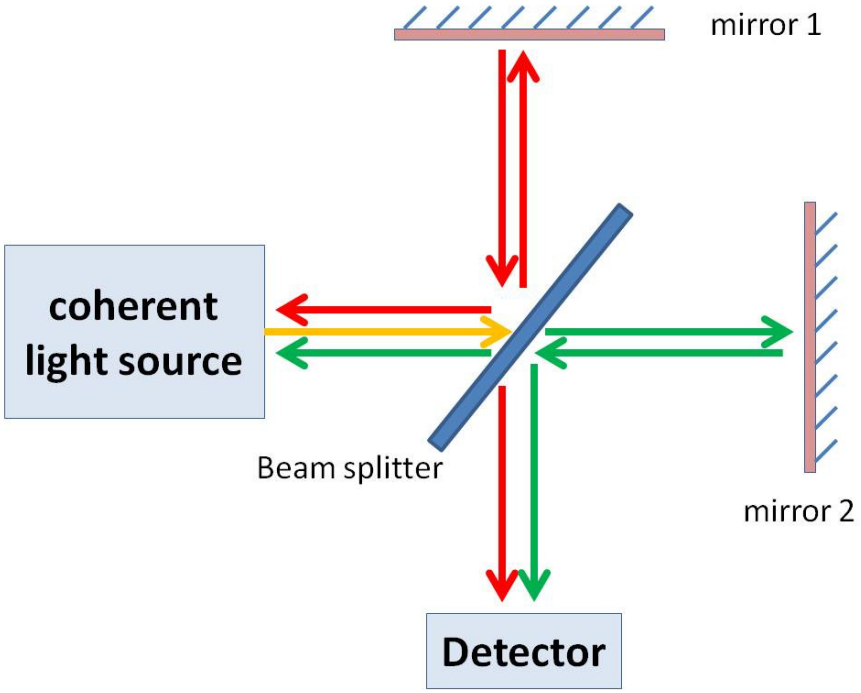

Figure 2.1: Michelson interferometer.

is $\tau = d/v_c$. Usually, only the intensity of the interference of those two lights are measured at the detector, whose value can be written as:

$$\begin{aligned} I &= \langle |A_1(t) + A_2(t + d/v_c)| \rangle \\ &\propto \langle |A_1| \rangle + \langle |A_2| \rangle + \langle A_1^* A_2 \rangle + \langle A_2^* A_1 \rangle \\ &= I_1 + I_2 + 2|A_1||A_2|cos(kd) \\ &= I_1 + I_2 + 2\mathrm{Re}\{C_t(\tau)\} \end{aligned} \tag{2.8}$$

k is the wavenumber of the light. I_1 and I_2 are the intensity of the light travels to mirror 1 and mirror 2, respectively. The signal obtained at the detector is proportional to the OPL difference. Thus, if one mirror is scanned, the signal at the detector will

be a function of time.

In addition to Michelson interferometer, other widely used interferometers include Mach-Zehnder interferometer, Fizeau interferometer and Sagnac interferometer, etc. The schematic configuration of Mach-Zehnder interferometer is shown in Fig. 2.2. It is also based on the OPL difference. Unlike Michelson interferometer, the light in the Mach-Zehnder interferomter only travel their its corresponding optical path once.

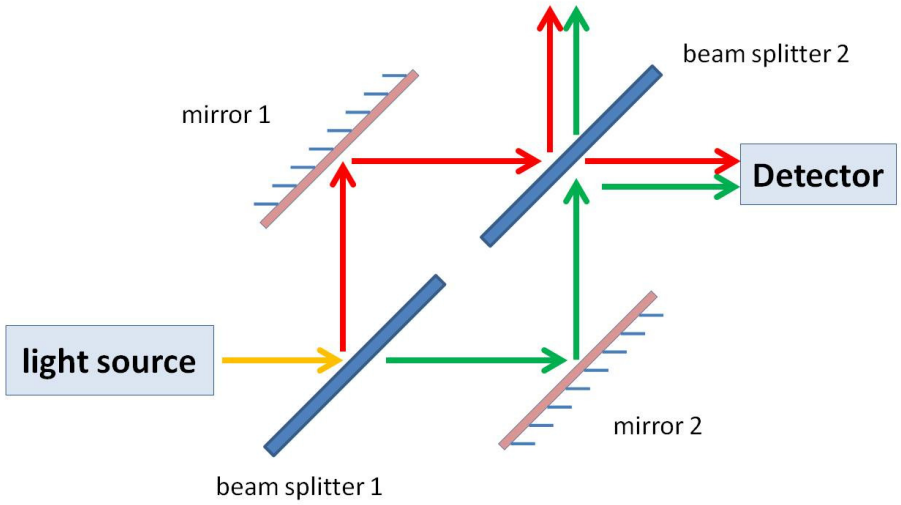

Figure 2.2: Mach-Zehnder interferometer.

A schematic configuration of Fizeau interferometer is shown in Fig. 2.3, which is widely used in the measurement of optical surfaces. The light reflected from the rear surface of the fiber interferences with the light reflected from the surface of the sample. One important property of the Fizeau interferometer is that two interference lights go through the same optical path. It eliminates the frequency mismatch between different optical paths, which causes dispersion for the Michelson interferometer. This property will be described in detail later. The Fizeau interferometer is the foundation

of common-path OCT (CP OCT).

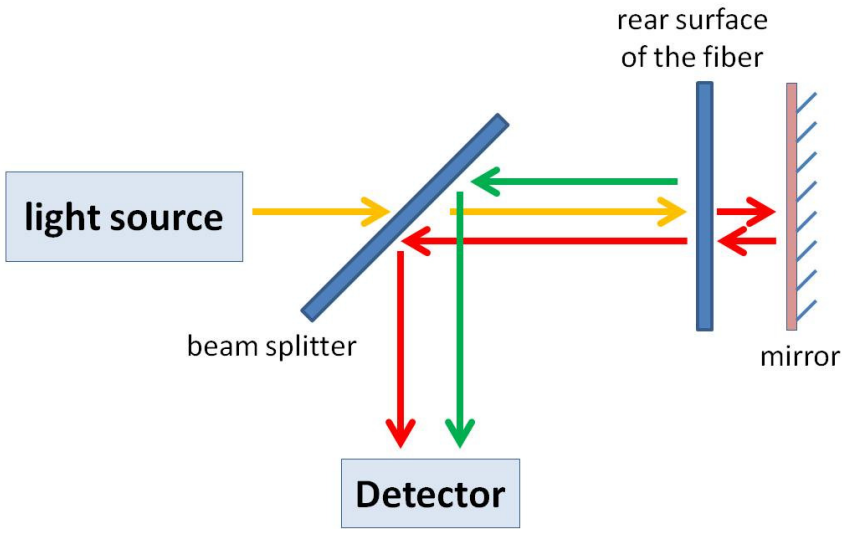

Figure 2.3: Fizeau interferometer.

Another type of interferometer, Sagnac interferometer, can be used for rotation sensing, such as gyroscope. A schematic configuration of Sagnac interferomter is shown in Fig. 2.4. A beam of light is splitted to two beams which counter-propagates around the same optical path. The interference of these two beams is detected after passing through the beam splitter again. The phase difference of these two beams is determined by the angular velocity of the apparatus.

2.3 Principles of OCT

In this section, Michelson interferometer based TD OCT and FD OCT systems are described. A schematic configuration of the Michelson interferometer-based OCT system is shown in Fig. 2.5. A sample is modeled as a multi-layer structure. Each

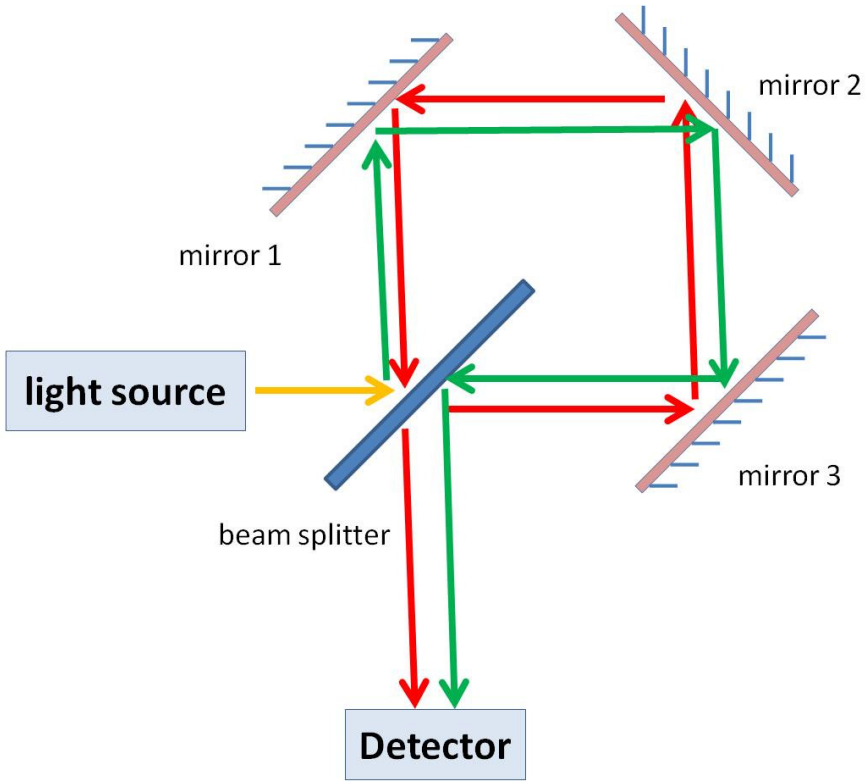

Figure 2.4: Sagnac interferometer.

layer is assumed to be a uniform media with a same refractive index. The light emitted from the light source is splitted at the beam splitter into two arms: the sample arm and the reference arm. Here, we assume a 50/50 beam splitter, which divides the income beam light in equal power. In practical applications, using an unbalanced beam splitter has many advantage. The light in the reference arm is incident upon a mirror and is reflected back to the beam splitter. The light in the sample arm is focused onto the multi-layer sample. The back-reflected light from the sample passes through the sample arm and is combined with the back-reflected reference light. The interference of these two lights is detected at the detector, which is processed to show

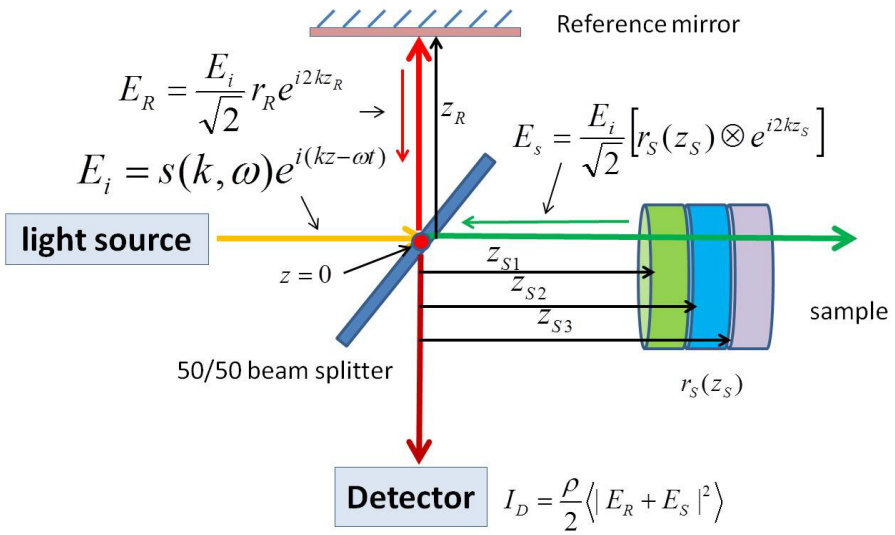

Figure 2.5: Schematic of the Michelson interferometer-based OCT system.

the depth profile of the sample around the focusing point (A-scan). As the focused point is swept across the sample surface, the cross-sectional imaging of the sample can be obtained by aligning the A-scans from different lateral positions, which is called a B-scan. The volumetric imaging of the sample (C-scan) can be achieved by obtaining multiple B-scans laterally. In this image acquisition scheme, the acquisition and reconstruction of different A-scans in the B-scan or C-scan are independent to each other, which plays an important role in CS SD OCT. The A, B, C scan of an OCT system is demonstrated at Fig. 2.6.

The electronic field of the light from the light source in complex form can be denoted as:

$$E_i = s(k,\omega)e^{i(kz-\omega t)} \tag{2.9}$$

$s(k,\omega)$ is the amplitude, which is a function of k and ω. $k = 2\pi/\lambda$ is the wavenumber.

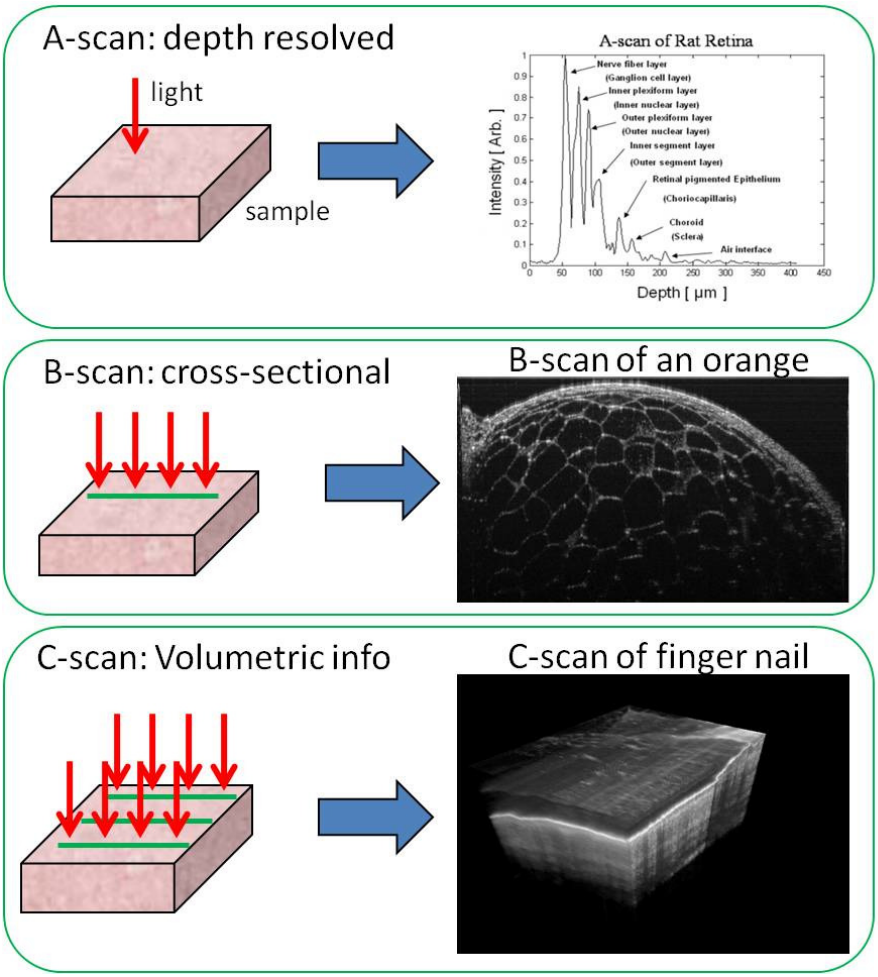

Figure 2.6: Demonstration of A-scan, B-scan, and C-scan of an OCT system.

λ is the wavelength. $\omega = 2\pi\nu$ is the angular frequency. ν is the frequency.

The middle point of the beam splitter is denoted as $z = 0$. The distance from the beam splitter to the reference mirror is represented as z_R. The reflector is characterized by its reflectivity r_R. Then the reflected light returned to the beam splitter in the reference arm is

$$E_R = \frac{E_i}{\sqrt{2}} r_R e^{i2kz_R} \tag{2.10}$$

Similarly, characterize the multi-layer sample by the reflectivity r_S of each layer and its distance from the beam splitter z_S:

$$r_S(z_S) = \sum_{n=1}^{N} r_{S_n} \delta(z_S - z_{S_n}) \tag{2.11}$$

N is the number of layers. Then the light returned to the beam splitter from the sample arm is

$$E_S = \frac{E_i}{\sqrt{2}} \left[r_S(z_S) \otimes e^{i2kz_S} \right] \tag{2.12}$$

The light power from both arms is reduced to 50% after traversing the beam splitter again and interferences at the detector. The photo current generated at the detector is the integral of the square of the electronic field, which equals the sum of these two light waves:

$$I_D(k,\omega) = \frac{\rho}{2} \langle |E_R + E_S|^2 \rangle = \frac{\rho}{2} \langle (E_R + E_S)^* (E_R + E_S) \rangle \tag{2.13}$$

ρ is the responsivity of the detector with the unit Amperes/Watt.

Substitute E_R and E_S with their forms in Eq. (2.10) and Eq. (2.12). Equation (2.13) can be rewritten as:

$$I_D(k,\omega) = \frac{\rho}{2} \left\langle \left| \frac{s(k,\omega)}{\sqrt{2}} r_R e^{i(2kz_R - \omega t)} + \frac{s(k,\omega)}{\sqrt{2}} \sum_{n=1}^{N} e^{i(2kz_{S_n} - \omega t)} \right| \right\rangle \tag{2.14}$$

In all applications, the response time of the detector is always much longer than

the angular frequency ω. Expanding Eq. (2.14) and removing the terms containing the angular frequency, it can be shown that Eq. (2.14) can be written as:

$$\begin{aligned} I_D(k) = \; & \frac{\rho}{4}\left[S(k)(R_R) + \sum_{n=1}^{N} R_{S_n}\right] \\ & + \frac{\rho}{8}\left[S(k)\sum_{n=1}^{N}\sqrt{R_R R_{S_n}}\cos(2k(z_R - z_{S_n}))\right] \\ & + \frac{\rho}{8}\left[S(k)\sum_{m\neq n=1}^{N}\sqrt{R_{S_m} R_{S_n}}\cos(2k(z_{S_m} - z_{S_n}))\right] \end{aligned} \tag{2.15}$$

$S(k) = \langle |s(|k,\omega)|^2 \rangle$, which is the power spectrum of the light source. $R_R = |r_R|^2$ is the power reflectivity of the reference mirror. R_{S_n} is the power reflectivity of the n-th layer of the sample.

2.3.1 Time Domain OCT

The first TD OCT system based on the low-coherent Michelson interferometer was introduced by David Huang et al in 1991 [2]. The schematic configuration of the fibre optics based TD OCT is shown in Fig. 2.7. TD OCT uses a broadband light source to illuminate the interferometer and a single detector to capture the interferometric signal. In TD OCT, the OCT depth scan is realized by moving the reference mirror mechanically in the axial direction, which limits the imaging speed of TD OCT. In the last decade, TD OCT has been replaced by FD OCT which eliminates the need for a mechanical axial scan and exhibits improvements in both imaging speed and sensitivity.

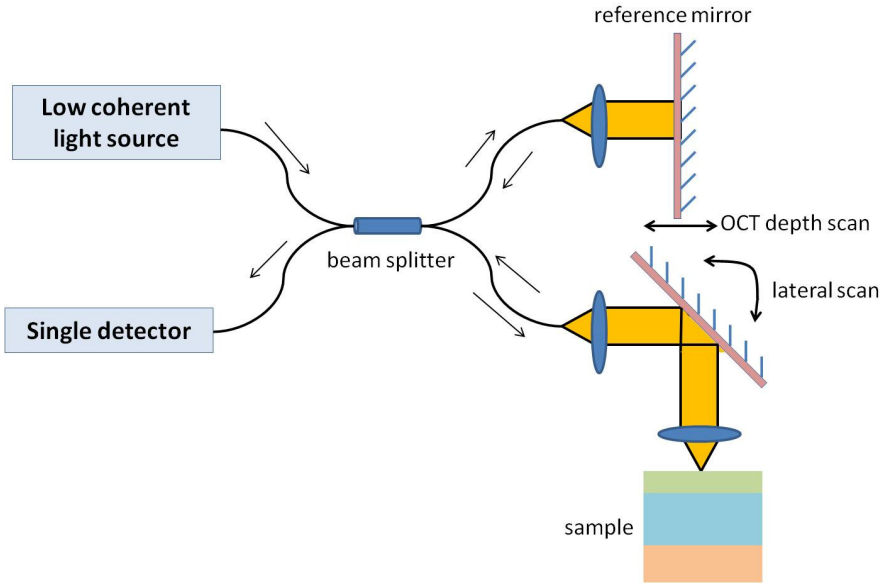

Figure 2.7: Schematic of TD OCT.

2.3.2 Fourier Domain OCT

The Fourier domain OCT (FD OCT) was first described by Fercher et al at 1995 [8]. In the last decade, it has been developed rapidly and now most of OCT systems are of this type. Compared to TD OCT, SD OCT has more than two orders of magnitude higher sensitivity and much faster imaging speed which recently has achieved more than 1 MHz A-scan rate.

As demonstrated above, there are mainly two types of configurations for FD OCT: spectral domain OCT (SD OCT) which uses a broadband light source and a dispersive spectrometer with a line-scan array detector, and swept source OCT (SS OCT) which uses a narrow-band swept-laser light source and a point detector. The works in this dissertation focuses on the development of CS SD OCT. However, since SD OCT and SS OCT are principally quite similar with a similar detector output, most of the work

in this dissertation can be easily adapted to the application of CS SS OCT.

Figure 2.8 schematically shows the layout and signal processing steps of SD OCT. Compared to TD OCT, SD OCT uses a fixed reference mirror in the reference arm.

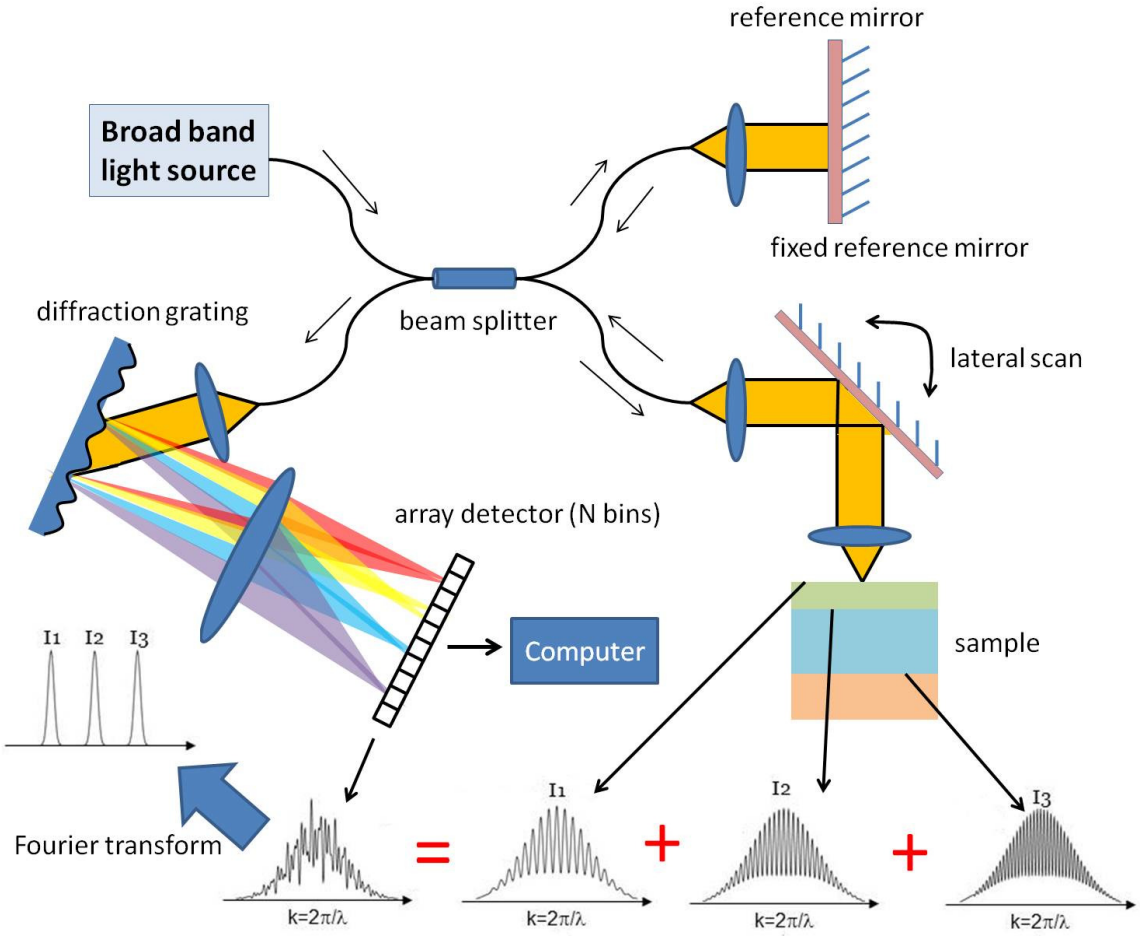

Figure 2.8: Schematic configuration and signal processing of SD OCT.

The spectrometer in SD OCT contains a diffraction grating that disperse the broad-band light, several collimating lens and a CCD or CMOS array detector. The signal arrives at the detector is the combined interferogram of the light waves from different depth within the sample. The depth profile of the sample at the focal point can be

obtained by taking the Fourier transform of the spectrum in Eq. (2.15). The result in spatial domain can be expressed as [3]:

$$\begin{aligned} i_D(z) = \; & \frac{\rho}{8}\left[\gamma(z)(R_R + \sum_{n=1}^{N} R_{S_n})\right] && \text{DC terms} \\ & + \frac{\rho}{8}\left[\gamma(z) \otimes \sum_{n=1}^{N} \sqrt{R_R R_{S_n}}\delta(z \pm 2(z_R - z_{S_n}))\right] && \text{cross-correlation terms} \\ & + \frac{\rho}{8}\left[\gamma(z) \otimes \sum_{m \neq n=1}^{N} \sqrt{R_{S_m} R_{S_n}}\delta(z \pm 2(z_{S_m} - z_{S_n}))\right] && \text{auto-correlation terms} \end{aligned} \tag{2.16}$$

where $\gamma(z)$ is the Fourier transform of $S(k)$. The "DC terms" corresponds to the spectrum of the light source. Usually, this is the largest component of the detector signal, which needs to be subtracted from the A-scan before displaying. The "cross-correlation terms" is the desired component of SD OCT. It contains several peaks whose locations are determined by the distance difference between the reference mirror position z_R and the position of the reflector z_{S_n}. The amplitude of these peaks changes according to the light power and the reflectivity of the reference and the media at position z_{S_n} within the sample. The last component, "auto-correlation terms", comes from the interference of the light between different reflectors. It is a source of the ghost artifact in regular SD OCT system. However, this component is usually located away from the desired signal, since the distances between the different reflectors in the sample are small.

In most SD-OCT system, the spectrum in Eq. (2.15) is sampled uniformly in wavelength, which implies they are nonlinear in wavenumber. Thus, applying the

discrete Fourier transform or fast Fourier transform to such signal will seriously degrade the image quality. Special procedure, both in hardware and software, has been developed to reconstruct the image from the non-linear wavenumber spectrum. Compared to the hardware solutions that usually complicate the design of the spectrometer and increase the cost, the software solutions are usually much more flexible and cost-efficient. There are two widely used software methods: the first one is based on the numerical interpolation that includes various linear interpolations and cubic interpolation; the other one uses the non-uniform discrete Fourier transform or the non-uniform fast Fourier transformation. These two methods will be demonstrated in detail later.

2.3.3 Axial resolution of SD OCT

If assuming a light source have a Gaussian shape spectrum with a bandwidth $\Delta\lambda$ in wavelength and Δk in wavenumber, it can be described mathematically:

$$S(k) = \frac{1}{\Delta k\sqrt{\pi}} e^{-\left[\frac{(k-k_0)}{\Delta k}\right]^2} \tag{2.17}$$

Here, k_0 is the central wavenumber.

It can be shown that its Fourier transform $\gamma(z)$ is:

$$\gamma(z) = e^{-z^2\Delta k^2} \tag{2.18}$$

From Eq. (2.16), the A-scan signal is the convolution of $\gamma(z)$ and the samples structure function $\delta(z \pm 2(z_R - z_{S_n}))$. Thus, the resolution of SD OCT in axial direction is defined as the full width at half maximum (FWHM) of $\gamma(z)$:

$$l_{axial} = \frac{2\sqrt{\ln(2)}}{\Delta k} = \frac{2\ln(2)}{\pi}\frac{\lambda_0^2}{\Delta\lambda} \tag{2.19}$$

Here, λ_0 is the central wavenumber of the light source. The axial resolution is determined by the bandwidth of the light source. Thus a broad band light source is usually used in SD OCT system.

2.3.4 Lateral resolution of SD OCT

In SD OCT, the lateral resolution is defined as the FWHM of the amplitude distribution at the beam waist of the probe beam. Assume the numerical aperture of the objective lens before the sample is denoted as NA. Then the lateral resolution of SD OCT can be expressed as:

$$l_{lateral} = \frac{2\sqrt{\ln(2)}}{\pi}\frac{\lambda_0}{NA} \tag{2.20}$$

2.3.5 Imaging depth of SD OCT

In SD OCT, the imaging depth is influenced by two factors. The first one is the sample's scattering and absorption effect. The intensity of the light will decrease as

the depth increases. Scattering changes the light direction and make it less likely to be detected by the detector. Even if detected, the light may have lost the coherence; thus will not contribute to the signal intensity. Absorption reduces the intensity of the light directly. For the near infrared wavelength light, scattering loss is larger, which has more effect on the signal level than absorption. Another factor is the spectral resolution of the spectrometer. It is determined by the coverage of the wavenumber Δk and the number of the pixels of the detector, which is denoted as N. According to the Shannon/Nyquist theory, the maximum imaging depth of SD OCT system is given by:

$$z_{max} = \frac{N\pi}{2\Delta k} \tag{2.21}$$

It can be figured out from Eq. (2.21) that the number of the pixels of the detector is determined by two components: the maximum imaging depth and the bandwidth of the light source:

$$N = \frac{2z_{max}\Delta k}{\pi} \tag{2.22}$$

Equation (2.19) shows that the axial resolution of SD OCT is inverse proportional to the bandwidth of the light source. Thus, a large number of measurements are needed for SD OCT imaging that requires both large imaging depth and high axial resolution. This requires a large linear array camera which is usually expensive. In addition, a large sampling rate will increase the imaging time, which makes the imaging susceptible to the motion artifact. It also becomes a heavy burden on the

image storage and transferring. The focus of this dissertation is to alleviate these requirement on the sampling rate, as well as the measurement size by implementing compressive sensing (CS) in SD OCT.

2.3.6 Sensitivity of SD OCT

The sensitivity of an SD OCT system can be expressed as [11]:

$$\sum_{SDOCT} = \frac{\frac{1}{N}\left(\frac{\rho\eta T}{h\nu_0}P_0\right)^2 \gamma_r\gamma_s R_r}{\frac{\rho\eta T}{h\nu_0}\frac{P_0}{N}\gamma_r R_r\left[1+\frac{(1+\Pi^2)}{2}\frac{\rho\eta}{h\nu_0}\frac{P_0}{N}\gamma_r R_r\frac{N}{\Delta\nu_{eff}}\right]+\sigma_{rec}^2} \tag{2.23}$$

Here, N is the number of pixels obtained at the detector. ρ is the efficiency of the spectrometer. η denotes the quantum efficiency of the detector. T is the CCD/CMOS detector integration time. h is Planck's constant. ν_0 is the central frequency. P_0 is the output of the source power. γ_r and γ_s are the parts of the input power that will exit the interferometer from the reference and sample arms, respectively. R_r is the power reflectivity of the reference mirror. Π is the polarization of the source. $\Delta\nu_{eff}$ is the effective spectral line width of the light source. σ_{rec} is the RMS of the receiver noise. The three terms in the denominator of Eq. (2.23) have different meanings: the first one represents the shot noise; the second one represents the excess noise; the third one is the receiver noise.

2.4 Dispersion and compensation in SD OCT

Dispersion is a common phenomenon for all the Michelson-based FD OCT systems, which degrades the axial resolution of the image and reduces the sensitivity of the system. The main reason for the dispersion is the presences of unbalanced components between the sample arm and the reference arm. It can be characterized by the frequency-dependent propagation constant $\beta(\omega)$ of the material. Expanding $\beta(\omega)$ near the central angular frequency ω_0 as a Taylor series leads to:

$$\beta(\omega) = \beta(\omega_0) + \left.\frac{d\beta}{d\omega}\right|_{\omega_0} (\omega - \omega_0) + \frac{1}{2}\left.\frac{d^2\beta}{d\omega^2}\right|_{\omega_0} (\omega - \omega_0)^2 + \frac{1}{6}\left.\frac{d^3\beta}{d\omega^3}\right|_{\omega_0} (\omega - \omega_0)^3 + \dots \quad (2.24)$$

The first term $\beta(\omega_0)$ is a constant. The second term $d\beta/d\omega$ is the inverse group velocity. The third term $d^2\beta/d\omega^2$ is the group velocity dispersion (GVD). It causes the broadening of the PSF and is the main source of dispersion in SD OCT. The fourth term, known as third-order dispersion term,$d^3\beta/d\omega^3$ produces the asymmetric pulse distortion of the PSF.

Dispersion compensation methods have been successfully implemented using both hardware and software approaches. Compare to the hardware methods which involve physically matching the dispersion of the reference and the sample arms, numerical dispersion compensation methods can be more flexible and cost-effective. One widely

used method is proposed in [1] by first resampling the non-linear wavenumber spectra with numerical interpolation; then constructing the complex representation of the real spectra with Hilbert transform; finally correcting the dispersion phase of the linear wavenumber complex signal. Compensating the second and third order dispersion is usually sufficient which implies the correcting phase is $\Phi(\omega) = -a_2(\omega - \omega_0)^2 - a_3(\omega - \omega_0)^3$, where a_2 and a_3 are constant. These two coefficients are usually set empirically by the user. Nevertheless, they can also be obtained automatically by optimizing the sharpness of the OCT image with an iterative procedure [1].

Chapter 3

Introduction to Compressive Sensing

Traditional signal sampling and reconstruction is based on the principle of well-known Shannon/Nyquist theory, which states that the signal can be reconstructed exactly from a set of measurements whose sampling rate is twice of the highest frequency of the signal. This principle is the foundation of digital signal processing and has been applied in many fields, including electronics, audios, and videos, and medical imaging.

However, as the bandwidth demand for electronic signals grows, the Shannon/Nyquist sampling rate requirement becomes a too large of demands in many fields. In general, a large sampling rate increases the overall system cost, requires a longer data acquisition time and puts a huge demand on the communication bandwidth and storage

resources. This situation is especially serious in many medical imaging modalities. Long acquisition time increases patients' radiation exposure time as well as discomfort and makes the imaging susceptible to unavoidable motion artifact.

Surprisingly, while the dimension of the medical imaging data is large in many cases, the relevant information resides in a lower dimension space. This led to a use of data compression schemes, which is widely used to reduce the amount of data needed for transmission and storage. Data compression tries to find a transformation domain that provides a sparse or compressible representation of the signal. In the transformation domain, the signal can be well-approximated by a few non-zero coefficients whose number is much smaller than the signal's size. Data compression is embedded in many popular image formats, such as JPEG, JPEG2000, MPEG and MP3.

Over the last decade, compressive sensing [21,22] has been rigorously studied as a revolutionary new approach to acquire and reconstruct a signal at a rate significantly lower than the Nyquist rate, provided that the signal has a sparse representation. Rather than first sampling the signal at a high rate and then compressing the sampled data, CS directly captures the data in a lower sampling rate. In other words, CS combines the sampling and compression into one step by measuring minimum samples that contain maximum information about the signal. CS has received significant success in many fields, especially in its applications in the medical imaging. In this chapter, basic principles of CS will be introduced.

3.1 Sparsity

CS relies on the observation that natural signals, such as sound and image, can be well-approximated by just a portion of the coefficients of their projection in a well chosen basis. Mathematically, if a signal $\mathbf{x}$ has only k non-zero coefficients, $\mathbf{x}$ is said to be k-sparse. In other words, $\mathbf{x}$ has sparsity k. This can be expressed as:

$$\sum_k = \mathbf{x} : \|\mathbf{x}\|_0 \leq k \tag{3.1}$$

to denote the set of all k-sparse signals. $\|\cdot\|_0$ represents the l_0-norm. Usually, the signal is not sparse itself; but is sparse in some basis Φ. In this case, we still refer $\mathbf{x}$ as k-sparse, with the understanding that $\mathbf{x}$ can be expressed as $\mathbf{x} = \Phi\mathbf{c}$, where $\|\mathbf{c}\|_0 \leq k$. However, from the application point of view, no signal in reality is perfectly k-sparse. We alleviate the definition: a signal is k-sparse if all but k of its coefficients are below a threshold τ. As τ goes to zero, this definition is the same as the k-sparse definition.

Widely used compressive-type basis includes wavelet transformation, curvelet transformation, localized sinusoids and fractal-type waveforms.

3.2 Acquisition Model

The signal acquisition model of CS is quite similar to that of traditional sensing frame work. If $\mathbf{x} \in R^N$ is the signal of interest, the sensing process can be expressed

as:

$$\mathbf{y} = \Phi\mathbf{x} \tag{3.2}$$

Here, $\mathbf{y}$ is the measurement vector. Φ is a M-by-N sensing matrix. Under the traditional sensing paradigm, M has to be equal to N for an exact or accurate reconstruction of $\mathbf{x}$ from $\mathbf{y}$. However, the aim of CS is to reconstruct $\mathbf{x}$ with a high fidelity from a much smaller number of measurements which implies that M is far less than N, given that $\mathbf{x}$ itself is sparse in some transformation domain. Figure 3.1 shows schematically the sensing process of CS.

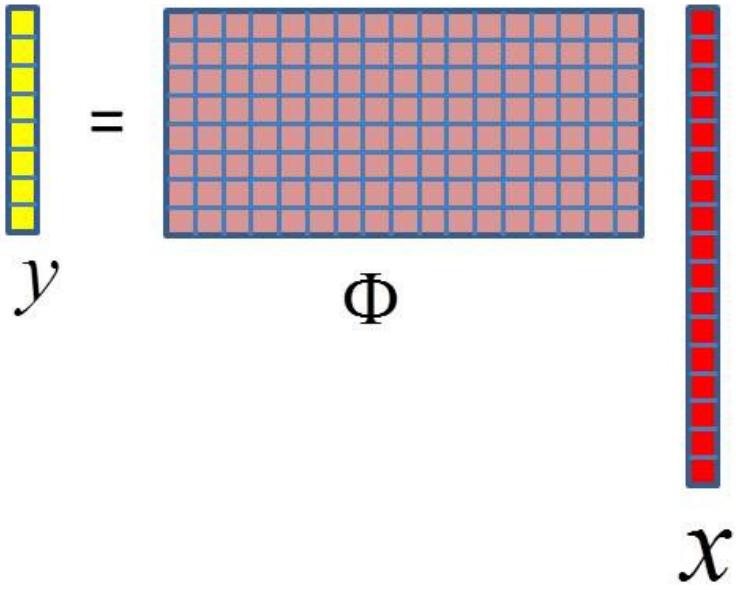

Figure 3.1: Schematic sensing process of CS.

Intuitively, the most straightforward strategy to reconstruct $\mathbf{x}$ from $\mathbf{y}$ is to search for the sparsest vector $\mathbf{x}$ among all candidate vectors that are consistent with the

observation vector $\mathbf{y}$, which is called l_0-minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \Phi\mathbf{x} \tag{3.3}$$

This problem, however, is highly non-convex and NP-hard in general [61, 62], which implies that the algorithm used to solve the l_0-minimization problem is computationally intractable. Therefore, it is recasted as a convex optimization problem: l_1-minimization, whose solution coincides with the solution of l_0-minization problem [21, 22]:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \Phi\mathbf{x} \tag{3.4}$$

where $\|\cdot\|_1$ is the l_1-norm which is defined as $\|\mathbf{x}\|_1 = \sum_i |x_i|$. An example of l_1-minimization problem is shown in Fig. 3.2(a) to give an intuitive explanation as to why the solution of l_1-minimization coincides with the solution of l_0-minimization. $N = 2$ and $M = 1$. Then all the candidate solutions form a line: $\{\mathbf{z} : \Phi\mathbf{z} = \mathbf{y}\}$. The unit sphere of the l_1-norm of a vector in R^2 is defined as $\{\mathbf{x} : \|x\|_1 = a\}$ where a is a constant. It can be seen that there is a unique solution of the l_1-minimization problem, except for the cases that Φ is parallel to one of the lines in the l_1-norm unit sphere. In addition, the solution has maximum sparsity, i.e. only one nonzero entry.

A highly related problem is called l_2 minimization problem which is defined as:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_2 \quad \text{subject to} \quad \mathbf{y} = \Phi\mathbf{x} \tag{3.5}$$

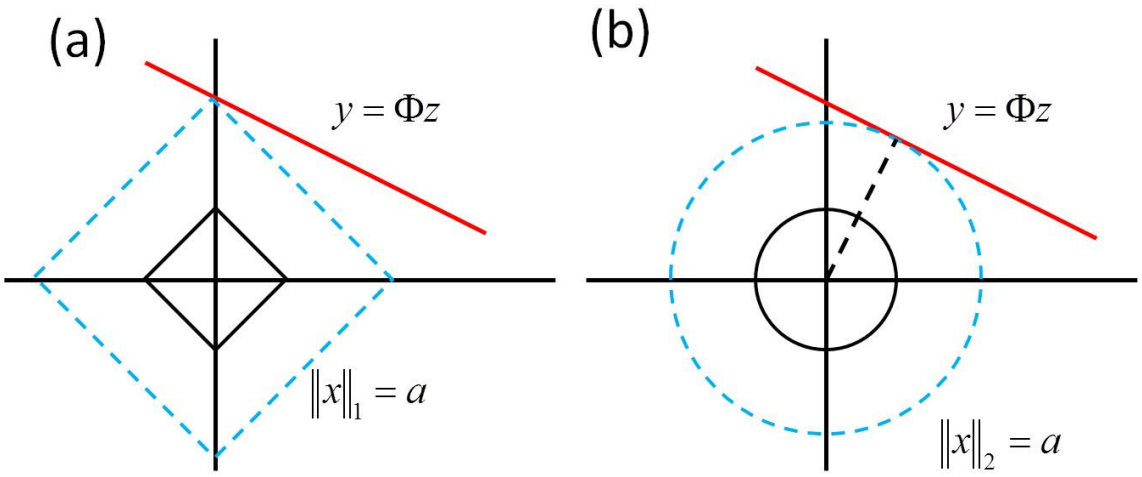

Figure 3.2: (a) a 2D example which demonstrates that the solution of l_1-minimization coincides with a sparest solution; (b) a 2D example which demonstrates that the solution of l_2-minimization is not sparse.

where $\|\mathbf{x}\|_2$ is the l_2-norm which is defined as $\|\mathbf{x}\|_2 = \sqrt{\sum_i |x_i|^2}$. The algorithm for solving the l_2-minimization problem is known as "linear regression", which gives the solution:

$$\mathbf{x} = \Phi^T(\Phi\Phi^T)^{-1}\mathbf{y} \tag{3.6}$$

Here, Φ^T is the transpose of Φ and $(\Phi\Phi^T)^{-1}$ denotes the inverse matrix of $\Phi\Phi^T$. Figure 3.2(b) shows a 2D example that demonstrates that the solution of l_2-minimization problem is not sparse.

3.3 Restricted Isometry Property

Figure 3.2(a) shows that not every Φ can be used as the sensing matrix. A sufficient condition for the sensing matrix Φ for an accurate reconstruction is described

in [63], which is named as Restricted Isometry Property (RIP). RIP states that if all subsets of s or fewer columns of Φ are nearly orthogonal, the linear program in Eq. (3.4) can obtain an exact or accurate enough reconstruction of $\mathbf{x}$. Its mathematical definition is given as:

Definition 3.1 (**Restricted Isometry Property (RIP)** [63])**.** *A matrix* $\Phi \in R^{M\times N}$ *satisfies RIP is there exists a constant* $\delta_s \in (0,1)$ *such that*

$$(1-\delta_s)\|x\|_2^2 \leq \|\Phi\mathbf{x}\|_2^2 \leq (1+\delta_s)\|x\|_2^2 \tag{3.7}$$

is true for every vector $\mathbf{x} \in R^N$ *with sparsity less than* s*.*

If $\delta_{2s} < 0.453$, the linear solution in Eq. (3.4) has a large possibility to reconstruct the s-sparse signal $\mathbf{x}$ stably.

A connection between RIP and CS is that if δ_{2s} is sufficiently small, all pairwise distances between any two s-sparse vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ will be preserved after the sensing process, which can be expressed mathematically as:

$$(1-\delta_{2s})\|\mathbf{x}_1-\mathbf{x}_2\|_2^2 \leq \|\Phi(\mathbf{x}_1-\mathbf{x}_2)\|_2^2 \leq (1+\delta_{2s})\|\mathbf{x}_1-\mathbf{x}_2\|_2^2 \tag{3.8}$$

Other definitions that are closely related to the construction of the sensing matrix include spark, mutual coherence, null-space property. The definition of spark is:

Definition 3.2 (**Spark**)**.** *The spark of a given matrix* Φ *is the smallest number of*

columns of Φ *that is linearly dependent.*

High spark implies that the sensing matrix contains more information, which improves the possibility for an accurate reconstruction of $\mathbf{x}$ in general.

The mutual coherence of a matrix measures the smallest angle between each pair of its columns.

Definition 3.3 (**Mutual Coherence).** *The mutual coherence of a matrix* $\Phi \in R^{M\times N}$ *is*

$$\mu(\Phi) = \max_{i\neq j} \frac{|\langle \phi_i, \phi_j \rangle|}{\|\phi_i\|^2 \|\phi_j\|^2} \tag{3.9}$$

The maximum mutual coherence of a matrix is 1 in the case that two columns are linearly dependent.

The null space property quantifies the notion that vectors in the null space of Φ should not be too concentrated on a small subset of indices:

Definition 3.4 (**Null Space Property (NSP)).** *A matrix* $\Phi \in R^{M\times N}$ *is said to satisfy the null space property (NSP) of order* s *if there exists a constant* $C > 0$ *such that*

$$\|\eta_T\|_1 \leq \gamma \|\eta_{T^c}\|_1 \tag{3.10}$$

for all sets $T \subset \{1, \ldots, N\}$, $\|T\|_0 \leq s$ *and for all* $\eta \in ker(\Phi)$

Here, η_T is the subvector that contains the components of η whose indices are the elements of T. T^c is the complementary set of T. $\ker(\Phi)$ is the null space of Φ which is defined as $\{\mathbf{x} : \Phi\mathbf{x} = 0\}$.

A matrix with high spark, low mutual coherence, and a small RIP constant is preferred during the design of the sensing matrix.

3.4 Randomness

However, there are a lot of problems when designing a sensing matrix that satisfies the properties mentioned above. Take RIP for instance. First, testing RIP for an arbitrary matrix is usually an intractable, combinatorial problem. Second, RIP provides guarantees over all signals. Its theory is necessarily limited by the worst case signals. Thus RIP is usually too strict for most applications, which increases the difficulty of finding a feasible solution.

Fortunately, RIP condition has been proven to hold for many types of random measurement matrices. As per available literatures, random Gaussian [64], random Bernoulli [65] and random Fourier or incoherent measurement matrix [21] satisfy RIP with high probability, as long as the number of measurements M obeys the following condition [22]:

$$M \geqslant s \log(N/s) \tag{3.11}$$

On the other hand, the sub-orthogonal matrix Φ whose rows are sampled uniformly from an orthogonal matrix A satisfies RIP when M is the order of $s \log^4(N)$.

3.5 CS Reconstruction

In practice, the measurements are likely to be contaminated by some form of noise, which can be expressed as:

$$\mathbf{y} = \Phi\mathbf{x} + \omega \tag{3.12}$$

where $\omega \in R^M$ is the noise signal. Assume that the energy bound of noise is $\|\omega\|_2 \leq \epsilon$. The l_1-minimzation problem in Eq. (3.4) is recasted as the basis pursuit denoisng (BPDN):

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Phi\mathbf{x}\|_2 \leq \epsilon \tag{3.13}$$

If Φ satisfies RIP and ϵ is not too big, the solution of Eq. (3.13) does not depart too far from the optimal solution [66, 67]. The error in the reconstruction proportionally grows with the increasingϵ.

The unconstrained version of BPDN is also widely used in the CS reconstruction algorithms which can be expressed as:

$$\min_{\mathbf{x}} \lambda\|\mathbf{x}\|_1 + \|\mathbf{y} - \Phi\mathbf{x}\|_2 \tag{3.14}$$

here, λ is the regularization parameter that determines the trade-off between the data consistency and sparsity.

There is another closed related and widely studied problem which is frequently

referred as the least absolute shrinkage and selection operator (LASSO) [68]:

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tau \tag{3.15}$$

Standard optimization theory [69] asserts that given properties ε, λ, and τ, Equations (3.13), (3.14), and (3.15) are equivalent.

A large number of efficient algorithms has been developed to solve CS problem, which can be roughly classified as either convex algorithms, greedy iterative algorithms, or combinationatorial/sublinear algorithms.

3.5.1 Convex Algorithms

This class of algorithms solves a convex optimization problem through linear programming. Convex optimization algorithms usually require a small number of measurements but are usually computationally complex. Equations (3.4), (3.13), (3.14), and (3.15) are all examples of convex problems. Developed convex algorithms include but are not limited to interior-point methods [22], projected gradient methods [70] and iterative thresholding [71], etc.

3.5.2 Greedy Algorithms

Greedy algorithms approximate the coefficients of the support of the signal which is implemented either by iteratively identifying the support of the signal until a con-

vergence criterion is met, or alternatively by obtaining an improved estimate of the signal at each iteration. Some greed algorithms have been proven to have performance guarantees that are very similar to the convex algorithms. The advantage of greedy algorithm is fast speed and easy to implement. The most well-known greedy algorithm is orthogonal matching pursuit (OMP) [72]. Other famous greedy algorithms include but are not limited to Stagewise OMP (StOMP) [73], Regularized OMP (ROMP) [74] and Compressive Sampling OMP (CoSaMP) [75].

3.5.3 Combinatorial/Sublinear Algorithms

Different from the convex algorithms and greedy algorithms, the combinatorial algorithms is mostly developed by the theoretical computer science community. These methods apply group testing to highly structured samples of the original signal, which is very fast and sublinear, but usually requires much more measurements than the algorithms in the other two previous cases.

Chapter 4

CS SD OCT with Dispersion Compensation on Non-linear Wavenumber Sampling

4.1 Introduction

Prior works on CS FDOCT required the spectral data in wavenumber to be linear for the under-sampling. This is because a uniform discrete Fourier transform (UDFT) matrix is used as the sensing matrix. Using UDFT is inherited from the conventional FDOCT image generation algorithms where FFT is used. However, in most FDOCT systems, the spectra are linear in wavelength and non-linear in wavenumber. Two methods have been proposed to obtain under-sampled linear wavenumber data from

the non-linear wavenumber spectra in SDOCT. The first method described in [30] uses a k-space grid filling method to remap the under-sampled non-linear wavenumber data to linearly-spaced wavenumber pixels based on a pre-calibrated wavenumber to pixel index. This method requires a spectral calibration with a numerical interpolation and is complicated and time-consuming. The second method uses a pre-calibrated k-linear random mask which enables acquiring a linear wavenumber subset by randomly under-sampling the non-linear wavenumber spectra directly [31]. The random mask is generated with the indices of the maximum and minimum points of the spectra by placing a single reflector in the sample arm. This method still needs a pre-calibration process. Also, even if a slight change in the sampling rate is desired, the whole calibration process needs to be repeated. Furthermore, it has an upper limit in the sampling rate because of the nature of the non-uniformity of the wavenumber.

Several approaches have been proposed to reconstruct OCT images from the non-linear in wavenumber spectra [76–82]. Non-uniform discrete or fast Fourier transform (NUDFT/NUFFT) is easy to implement and results in high quality images [79–82]. Applications of CS with the NUDFT/NUFFT matrix as the sensing matrix have also been studied, mainly for non-Cartesian sampling in magnetic resonance imaging (MRI) [83–86], which use a non-linear wavenumber spectral data directly. However, as will be shown in Sec. 4.3.1, CS with the under-sampled NUDFT/NUFFT matrix as the sensing matrix cannot be applied directly to the non-linear wavenumber under-sampling of FDOCT spectra because the reconstructed A-scan will have less sparsity

in spatial domain which requires more k-space sampling. Here the sparsity is defined as the number of coefficients to represent the signal is close to 0.

In this chapter, it is shown that modifying the NUDFT matrix by making them symmetric while preserving the intensity of the desired part of the A-scans can promote the sparsity of A-scans. Therefore, the modified NUDFT matrix can be used as the sensing matrix in CS reconstruction on the non-linear wavenumber sampling of FDOCT signal.

4.2 Mathematical background

4.2.1 Uniform discrete Fourier transform

In most FDOCT systems, the spectra are non-linear in wavenumber which requires pre-processing procedures such as numerical interpolation to convert the spectral dataset into linear in wavenumber if UDFT is to be applied. Denote $\mathbf{y} = [y_0, y_1, \ldots, y_{N-1}]^T$ as the non-linear wavenumber k-space spectra (real value) and $\hat{\mathbf{y}} = [\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_{N-1}]^T$ as the linear wavenumber spectral data (real value) obtained from $\mathbf{y}$. The uppercase T denotes the transpose. N is the whole signal length (no undersample). Denote $\mathbf{x} = [x_0, x_1, \ldots, x_{N-1}]^T$ as the A-scan reconstructed from $\mathbf{y}$ and $\hat{\mathbf{x}} = [\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_{N-1}]^T$ as the A-scan reconstructed from $\hat{\mathbf{y}}$. $\mathbf{k} = [k_0, k_1, \ldots, k_{N-1}]^T$ is the non-linear wavenumber corresponding to $\mathbf{y}$ while $\hat{\mathbf{k}} = [\hat{k}_0, \hat{k}_1, \ldots, \hat{k}_{N-1}]^T$ is the linear wavenumber corresponding to $\hat{\mathbf{y}}$. Then $\hat{\mathbf{x}}$ can be obtained from $\hat{\mathbf{y}}$ through

inverse UDFT:

$$\hat{x}_n = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \hat{y}_m \ \exp\left(i \frac{2\pi}{\Delta \hat{k}} (\hat{k}_m - \hat{k}_0) \times n\right) = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \hat{y}_m \ \exp\left(i \hat{\omega}_m \times n\right) \quad (4.1)$$

for $n \in [0, \ldots, N-1]$. i is the imaginary unit. $\Delta \hat{k} = \hat{k}_{N-1} - \hat{k}_0$ is the wavenumber range. $\hat{\omega}_m = 2\pi/N \times m$. The derivation of the last part of Eq. (4.1) is because $\hat{\mathbf{k}}$ contains linear wavenumber: $\Delta \hat{k}$ can be written as $\Delta \hat{k} = N \times \delta \hat{k}$ and $\hat{k}_m - \hat{k}_0 = m \times \delta \hat{k}$.

4.2.2 Non-uniform discrete Fourier transform

It has been shown that A-scan ($\mathbf{x}$) can be reconstructed from the non-linear wavenumber spectra ($\mathbf{y}$) through inverse NUDFT [79–82]. To avoid any numerical instability in computing the inverse NUDFT matrix, [82] uses the forward NUDFT matrix instead:

$$x_n = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} y_m \ \exp\left(i \frac{2\pi}{\Delta k} (k_m - k_0) \times n\right) = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} y_m \ \exp\left(i \omega_m \times n\right) \quad (4.2)$$

for $n \in [0, \ldots, N-1]$. $\Delta k = k_{N-1} - k_0$ is the wavenumber range ($\Delta k = \Delta \hat{k}$). $\omega_m = 2\pi/\Delta k \times (k_m - k_0)$. In standard FDOCT where $\mathbf{y}$ and $\hat{\mathbf{y}}$ are real values, only the first halves of $\mathbf{x}$ and $\hat{\mathbf{x}}$ are displayed, in other words, $n \in [0, \ldots, N/2 - 1]$ in Eqs. (4.1) and (4.2). Compared to the interpolation-UDFT method, inverse NUDFT method has several advantages: it is simple to implement and immune to interpolation

errors which result in increased background noise and side-lobes, especially at larger image depths [82].

In the rest of this chapter, if not specified, inverse NUDFT refers to a transformation with the inverse NUDFT matrix which has the same form as the forward NUDFT matrix in Eq. (4.2). Constructing a new sensing matrix by starting from the forward NUDFT matrix, instead of the strict inverse NUDFT matrix, will be discussed in Sec. 4.6.

4.2.3 Compressive sensing in FDOCT

In CS-FDOCT, A-scan can be reconstructed from an under-sampled spectral data by solving the following optimization problem as long as the data is sparse in some domain:

$$\underset{\mathbf{g}}{\text{minimize}} \ \|\mathbf{W}\mathbf{g}\|_1, \text{s.t.} \|\mathbf{F}_u\mathbf{g} - \mathbf{z}_u\|_2 \leqslant \varepsilon \tag{4.3}$$

where $\mathbf{g}$ is the desired A-scan signal in spatial domain. $\mathbf{W}$ is the sparsifying operator which will transform $\mathbf{g}$ to a sparse representation, such as the wavelet transform matrix. If $\mathbf{g}$ is sparse in spatial domain, $\mathbf{W}$ could be the identity matrix. $\mathbf{F}_u$ is the under-sampled sensing matrix and $\mathbf{z}_u$ is the under-sampled k-space data. ε dictates the fidelity of the reconstruction to the sampled data or, equivalently, it reflects the noise level of $\mathbf{z}_u$; $\varepsilon \approx 0$ for noise-free data.

If $\mathbf{F}_u$ is the under-sampled UDFT matrix, $\mathbf{z}_u$ has to be linear wavenumber sampled

data which is the case in traditional CS-FDOCT. However, as mentioned above, the linear wavenumber sampling from the non-linear wavenumber spectra of FDOCT requires either k-space grid filling with a spectral calibration or a k-linear random mask calibration.

If an under-sampled NUDFT matrix is used as the sensing matrix, $\mathbf{z}_u$ is no longer required to be linear in wavenumber and the non-linear wavenumber spectra of FDOCT can be under-sampled directly at an arbitrary sampling rate and used in CS reconstruction.

4.3 CS with the modified NUDFT matrix on the non-linear wavenumber sampling

4.3.1 A-scan sparsity of FDOCT using UDFT and NUDFT

CS reconstruction using the under-sampled non-linear wavenumber spectral data with a NUDFT matrix as the sensing matrix has already been successfully implemented, namely in the case of the non-Cartesian sampling of MRI [83–86]. Traditional CS-FDOCT on the UDFT matrix and the linear wavenumber sampling also

shows that A-scans of FDOCT are sparse enough and can be reconstructed from highly under-sampled data [30–32, 34–37, 87, 88].

However, applying the NUDFT matrix and non-linear wavenumber sampling to CS-FDOCT would be a problem since the sparsity of the reconstructed A-scan is much less than that of the A-scan obtained from traditional CS-FDOCT with the UDFT matrix and linear wavenumber sampling. According to CS theory, the sampling rate for an accurate reconstruction depends highly on the A-scan's sparsity. Decreased A-scan sparsity requires much more sampling to guarantee accruate image reconstruction.

The difference in sparsity of the A-scans with both methods is compared using the whole k-space spectra. If UDFT is applied, the A-scan ($\hat{\mathbf{x}}$) can be obtained from the linear wavenumber spectra ($\hat{\mathbf{y}}$) through Eq. (4.1). Because the elements of $\hat{\mathbf{y}}$ are real, $\hat{\mathbf{x}}$ is symmetric:

$$\begin{aligned} \hat{x}_{N-n} &= \tfrac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \hat{y}_m \ \exp\left(i\tfrac{2\pi}{N} m \times (N-n)\right) \\ &= \tfrac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \hat{y}_m \ \exp\left(-i\tfrac{2\pi}{N} m \times n\right) \\ &= (\hat{x}_n)^* \end{aligned} \tag{4.4}$$

for $n \in [1, \ldots, N/2-1]$ because m is integer and $\exp(i2\pi m) = 1$. $(\hat{x}_n)^*$ is the conjugate of $\hat{x}_n$. This conjugate property implies that the intensity of $\hat{x}_{N-n}$ is the same as that of $\hat{x}_n$ for $n \in [1, \ldots, N/2-1]$. Figure 4.1(a) shows the plot of an A-scan (belonging to a mouse paw scanning) obtained by applying inverse UDFT to $\hat{\mathbf{y}}$ which

is obtained from the non-linear wavenumber spectra ($\mathbf{y}$) using cubic interpolation.

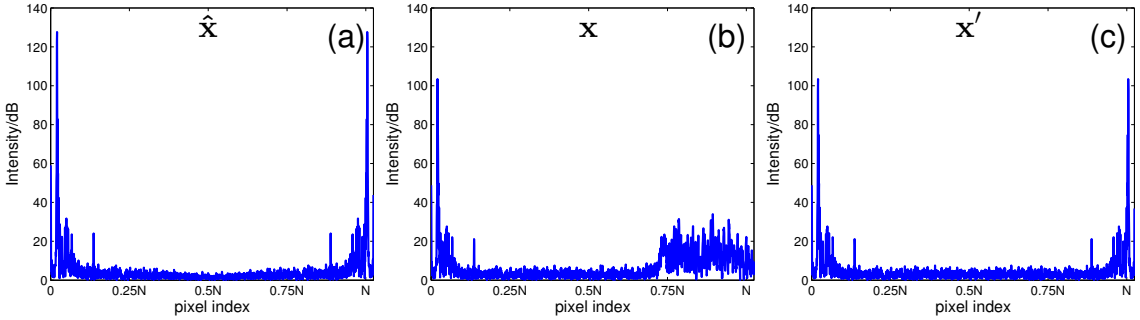

Figure 4.1: Sparsity comparison of A-scans by applying (a) inverse UDFT to the linear wavenumber whole spectra ($\hat{\mathbf{y}}$), (b) inverse NUDFT to the non-linear wavenumber whole spectra ($\mathbf{y}$), (c) modified inverse NUDFT to the non-linear wavenumber whole spectra ($\mathbf{y}$)

If NUDFT is applied, the A-scan ($\mathbf{x}$) can be obtained from the non-linear wavenumber spectra ($\mathbf{y}$) through Eq. (4.2). Because $\mathbf{y}$ is non-linear in wavenumber, $\mathbf{x}$ is no longer symmetric:

$$\begin{aligned} x_{N-n} &= \tfrac{1}{\sqrt{N}} \textstyle\sum_{m=0}^{N-1} y_m \ \exp\left(i\omega_m \times (N-n)\right) \\ &= \tfrac{1}{\sqrt{N}} \textstyle\sum_{m=0}^{N-1} y_m \ \exp\left(-i\omega_m \times n\right) \times \exp\left(i\omega_m N\right) \\ &\neq (x_n)^* \end{aligned} \tag{4.5}$$

for $n \in [1, \ldots, N/2-1]$ because usually $\exp(i\omega_m N) \neq \pm 1$. Figure 4.1(b) shows the plot of A-scan obtained by applying inverse NUDFT to $\mathbf{y}$.

As can be seen in Figs. 4.1(a) and 4.1(b), the first halves of $\hat{\mathbf{x}}$ and $\mathbf{x}$, i.e., $[\hat{x}_0, \ldots, \hat{x}_{N/2-1}]$ and $[x_0, \ldots, x_{N/2-1}]$ have similar sparsity. But the sparsity of the second half of $\mathbf{x}$ is much less than that of $\hat{\mathbf{x}}$. The decrease of sparsity implies that more sampling is required to reconstruct $\mathbf{x}$ than $\hat{\mathbf{x}}$ using CS.

Although a specific example from a mouse paw scanning is displayed in Fig. 4.1, ac-

cording to the experimental results with different samples, it is quite universal that $\mathbf{x}$ has less sparsity than $\hat{\mathbf{x}}$. The influence of the decreased sparsity usually cannot be eliminated by using a larger ε in CS reconstruction: In Fig. 4.1(b), $[x_{0.75N}, \ldots, x_{N-1}]$ has higher intensity than most of $[x_0, \ldots, x_{N/2-1}]$, which implies that most of $[x_0, \ldots, x_{N/2-1}]$ will receive a bigger penalty than $[x_{0.75N}, \ldots, x_{N-1}]$ during the reconstruction and is more likely to be zero with larger ε. Although the second halves of both $\mathbf{x}$ and $\hat{\mathbf{x}}$ will not be displayed for the standard FDOCT system, their sparsity will greatly influence the reconstruction of the FDOCT signal with CS since Eq. (4.3) tries to find the solution that minimizes the $l1$-norm of the whole A-scan.

4.3.2 The modified NUDFT (MNUDFT) matrix

The sparsity of the second half of $\mathbf{x}$ has a greater influence on the CS reconstruction of the whole $\mathbf{x}$. In standard FDOCT, however, the intensity of this undisplayed half of $\mathbf{x}$ is of no interest as long as its sparsity is high enough. It would be idea if the second half of $\mathbf{x}$ is always 0. However, that's not typically true for an arbitrary $\mathbf{y}$.

It has already been shown in previous work on CS-SDOCT that $\hat{\mathbf{x}}$ can be accurately reconstructed with a relatively small sample size [30–32]. Therefore our motivation is: since the first halves of $\mathbf{x}$ and $\hat{\mathbf{x}}$ have a similar sparsity and $\hat{\mathbf{x}}$ can be reconstructed by CS, if the second half of $\mathbf{x}$ is symmetric to its first half, $\mathbf{x}$ and $\hat{\mathbf{x}}$ will have a similar sparsity. Then $\mathbf{x}$ can be accurately reconstructed with almost the same amount of sampling required to reconstruct $\hat{\mathbf{x}}$ by CS. We find that this idea can be realized by modifying the NUDFT matrix.

The inverse NUDFT transformation in Eq. (4.2) can be written in matrix form as follows:

$$\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} h(0,0) & h(1,0) & \dots & h(N-1,0) \\ \vdots & \vdots & \ddots & \vdots \\ h(0,N/2) & h(1,N/2) & \dots & h(N-1,N/2) \\ \vdots & \vdots & \ddots & \vdots \\ h(0,N-1) & h(1,N-1) & \dots & h(N-1,N-1) \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} \tag{4.6}$$

where $h(p,q) = \exp(i\omega_p \times q)$.

It is easy to see that values of the bottom half of $\mathbf{x}$ ($[x_{N/2}, \dots, x_{N-1}]$) are only relevant to the bottom half of the inverse NUDFT matrix and $\mathbf{y}$. Therefore, obtaining symmetric reconstructed A-scan ($\mathbf{x'}$) can be realized by modifying the bottom half of the inverse NUDFT matrix:

$$\begin{bmatrix} x'_0 \\ x'_1 \\ \vdots \\ x'_{N-1} \end{bmatrix} = \begin{bmatrix} h(0,0) & h(1,0) & \dots & h(N-1,0) \\ h(0,1) & h(1,1) & \dots & h(N-1,1) \\ \vdots & \vdots & \ddots & \vdots \\ h(0,N/2-1) & h(1,N/2-1) & \dots & h(N-1,N/2-1) \\ h(0,N/2) & h(1,N/2) & \dots & h(N-1,N/2) \\ \mathbf{h(0,N/2-1)^*} & \mathbf{h(1,N/2-1)^*} & \dots & \mathbf{h(N-1,N/2-1)^*} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h(0,1)^*} & \mathbf{h(1,1)^*} & \dots & \mathbf{h(N-1,1)^*} \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} \tag{4.7}$$

where $h(p,q)^* = \exp(-i\omega_p \times q)$ is the conjugate of $h(p,q)$.

Elements from the second row to the last row of the bottom half of the transformation matrix (rows in bold) are conjugate to those of the symmetric rows in the top half, e.g., elements of the row corresponding to $x'_{(N/2+1)}$ are conjugate to the elements of the row corresponding to $x'_{(N/2-1)}$. It can be proved that x' is symmetric:

$$x'_{N-n} = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} y_m \; h(m,n)^* = \left(\frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} y_m \; h(m,n) \right)^* = (x'_n)^* \tag{4.8}$$

which implies that x'_{N-n} and x'_n have the same intensity for $n \in [1, \ldots, N/2-1]$. Thus $\mathbf{x}'$ is symmetric. It is also easy to see that $[x'_0, x'_1, \ldots, x'_{N/2-1}]$ is the same as $[x_0, x_1, \ldots, x_{N/2-1}]$. Thus, proposed modification preserves the intensity of the desired part of the A-scan. The first and $(N/2+1)th$ row of the modified inverse NUDFT matrix are unchanged, which makes $\mathbf{x}'$ and $\hat{\mathbf{x}}$ to have the same symmetric structure. Figure 4.1(c) shows the plot of $\mathbf{x}'$ whose sparsity is similar to $\hat{\mathbf{x}}$ and much higher than $\mathbf{x}$. Proposed method shows good performance at promoting the sparsity of A-scan.

The modified inverse NUDFT matrix in Eq. (4.7), however, cannot be used directly as the sensing matrix in CS-FDOCT because the sensing matrix should transform data from spatial domain to k-space according to Eq. (4.3). Thus, its inverse matrix, i.e. the modified NUDFT matrix is needed. Based on the fact that the inverse NUDFT matrix in Eq. (4.6) is indeed the forward NUDFT matrix instead of the strict inverse NUDFT matrix, the modified NUDFT matrix can be easily obtained by taking conjugate transpose of the transformation matrix in Eq. (4.7).

Using the under-sampled modified NUDFT matrix as the sensing matrix, A-scan can

be reconstructed with high accuracy by CS from under-sampled non-linear wavenumber spectral data.

4.4 CS with dispersion compensation

4.4.1 Dispersion compensation with NUDFT on non-linear wavenumber real spectra

Dispersion degrades FDOCT image quality by introducing a frequency-dependent phase term to the Fourier components of the signal. One widely used method to compensate the dispersion is proposed in [1] which first resamples the non-linear wavenumber spectra with numerical interpolation; generates the imaginary part of the signal with Hilbert transform; then corrects the phase of the linear wavenumber complex signal to compensate dispersion; finally applies UDFT to the corrected spectrum to obtain the A-scan. However, this method cannot be applied to under-sampled spectral data because both interpolation and Hilbert transform require whole spectra which makes it very difficult to obtain under-sampled dispersion compensated linear wavenumber spectral data. Also, this method cannot be applied as post processing to the reconstructed A-scan with dispersion because it corrects the dispersion in k-space, not in spatial domain. Thus it is difficult to incorporate this widely used dispersion compensation method into CS-FDOCT directly.

Therefore, we propose and validate a dispersion compensation method by first multiplying the correcting phase directly to the non-linear wavenumber real spectra; then applying

NUDFT to the corrected spectrum to reconstruct the A-scan. This method eliminates the need for interpolation and Hilbert transform which are used to transform the spectra to be first linear in wavenumber then complex. It will be shown in Sec. 4.4.3 that after transformation, proposed dispersion compensation method has the form similar to that of the inverse NUDFT-based FDOCT image generation algorithm in Eq. (4.6). Then the transformation matrix is modified in the same way mentioned in Sec. 4.3.2 to promote the A-scan's sparsity and its under-sampled matrix can be used in CS reconstruction on the non-linear wavenumber real sampling to obtain dispersion compensated A-scan.

Multiplying the correcting phase directly to the non-linear wavenumber real spectra can be written as:

$$\begin{aligned} I_{comp}(\omega_m) = &\quad Re\{2 \times \textstyle\sum_n \sqrt{S_n(\omega_m)S_r(\omega_m)}\exp(-i[\omega_m\tau_n + \Phi(\omega_m)])\} \times \exp(i\Phi(\omega_m)) \\ = &\quad (\textstyle\sum_n A_n(\omega_m)\,(\exp(-i[\omega_m\tau_n + \Phi(\omega_m)] + \exp(i[\omega_m\tau_n + \Phi(\omega_m)]))) \exp(i\Phi(\omega_m)) \\ = &\quad \underbrace{\textstyle\sum_n A_n(\omega_m)\exp(-i[\omega_m\tau_n])}_{A_1} + \underbrace{\textstyle\sum_n A_n(\omega_m)\exp(i[\omega_m\tau_n + 2\Phi(\omega_m)])}_{A_2} \end{aligned} \tag{4.9}$$

for $m \in [0, \ldots, N-1]$. $\mathbf{I}_{comp}(\omega)$ is the corrected spectrum. $\mathbf{S_n}(\omega)$ is the intensity of light reflected from the n-th layer in the sample; $\mathbf{S_r}(\omega)$ is the intensity of light reflected from the reference arm; τ_n is the optical group delay of the *n-th* reflection to the reference light path. $\Phi(\omega)$ is the dispersion term. $A_n(\omega_m)$ substitutes $\sqrt{S_n(\omega_m)S_r(\omega_m)}$ from the second row. The first term after the first equal mark is the interferometric signal with some degree of dispersion (i.e. $\mathbf{y}$ in previous sections) while the second term is the dispersion compensation term. Compensating second and third order dispersion is usually sufficient where $\Phi(\omega) = -a_2(\omega - \omega^*)^2 - a_3(\omega - \omega^*)^3$. ω^* is the central angular frequency; a_2 and a_3

are constant. In the last row, the term A_1 is the desired dispersion compensated spectra. However, proposed method will also introduce an undesired term in $\mathbf{I}_{comp}(\omega)$: A_2.

Then inverse NUDFT is applied to the corrected spectrum $\mathbf{I}_{comp}(\omega)$. Denote the results of applying the inverse NUDFT to A_1 and A_2 as T_1 and T_2 respectively. The resulting A-scan is the sum of T_1 and T_2. In standard FDOCT system, only the first half of the A-scan will be shown. The first half of T_1 is the desired dispersion compensated A-scan with better resolution while the first half of T_2 degrades the A-scan's resolution. However, as will be shown below, the intensity of first half of T_2 is relatively small compared to that of the first half of T_1. In other words, T_1 dominates the displayed half of the A-scan and the resolution degradation caused by T_2 has little effect.

4.4.2 Experimental validation

To illustrate this domination effect, the simulation with only one reflector ($n = 1$) is shown in Fig. 4.2 which plots the intensity ratio value of $(|T_1(\tau_1)|/|T_2(\tau_1)|)$ with different reflector position τ_1 (from 1 to $N/2 - 1$). The simulation is done with different level of dispersion ($a_2 \in \{-500, -250, -100, 0, 100, 250, 500\} fs^2$; $a_3 = 0$) to demonstrate that intensity of T_1 is much higher than that of T_2 regardless of the level of dispersion. ω is obtained from the SDOCT system used in the study. S_n and S_r are set as 1 since their values do not influence the plot.

As can be seen in Fig. 4.2, most of $(|T_1(\tau_1)|/|T_2(\tau_1)|)$ is above 30dB when the reflector is in the displayed half of the A-scan. T_1 dominates T_2 under various dispersion condition. The curve without dispersion shows higher ratio as expected.

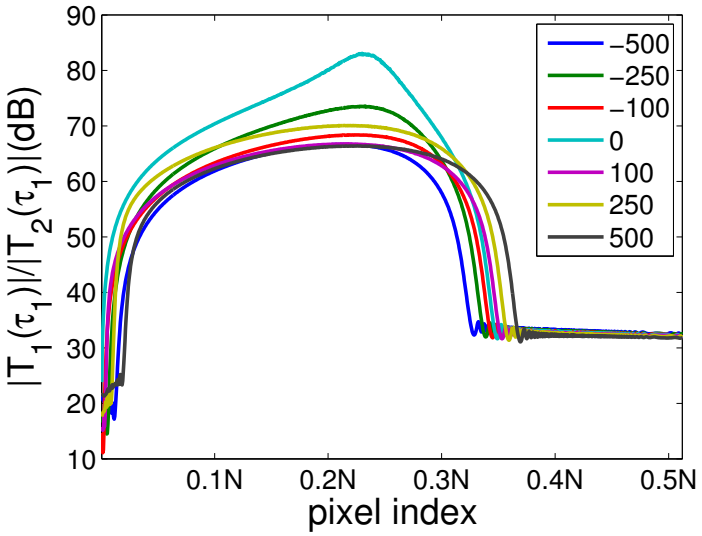

Figure 4.2: Plot of $|T_1(\tau_1)|/|T_2(\tau_1)|$ versus different reflector position τ_1. Simulation is done with different level of dispersion ($a_2 \in \{-500, -250, -100, 0, 100, 250, 500\} fs^2$; $a_3 = 0$).

Sensitivity roll-off of the proposed dispersion compensation method on the non-linear wavenumber real spectra ($\mathbf{y}$) is compared with those by applying (1) inverse UDFT to the linear wavenumber real spectra ($\hat{\mathbf{y}}$) without dispersion compensation; (2) inverse NUDFT to $\mathbf{y}$ without dispersion compensation; (3) dispersion compensation method in [1]. Here 124 A-scans with 2048-pixel each were averaged for each position. A 2cm water cell was inserted to the reference arm of the interferometer to introduce large dispersion mismatch (both $\mathbf{y}$ and $\hat{\mathbf{y}}$ contains dispersion). The dispersion coefficients were empirically set as $a_2 = 460 fs^2$ and $a_3 = 134 fs^3$. The results are shown in Fig. 4.3 (intensity of each A-scan is normalized). Figure 4.3(b) shows flatter sensitivity roll-off with smaller background noise and side-lobes at larger image depth than Fig. 4.3(a), which is consistent with the observation in [79, 82]. Peaks in Figs. 4.3(c) and 4.3(d) are much sharper than those in Figs. 4.3(a) and 4.3(b), indicating the FWHM of $4\mu m$ against $22\mu m$. Figure 4.3(d) also

shows better sensitivity than Fig. 4.3(c) as well as less background noise and side-lobes. Thus, proposed method shows good potential at compensating dispersion and outperforms the dispersion compensation method in [1].

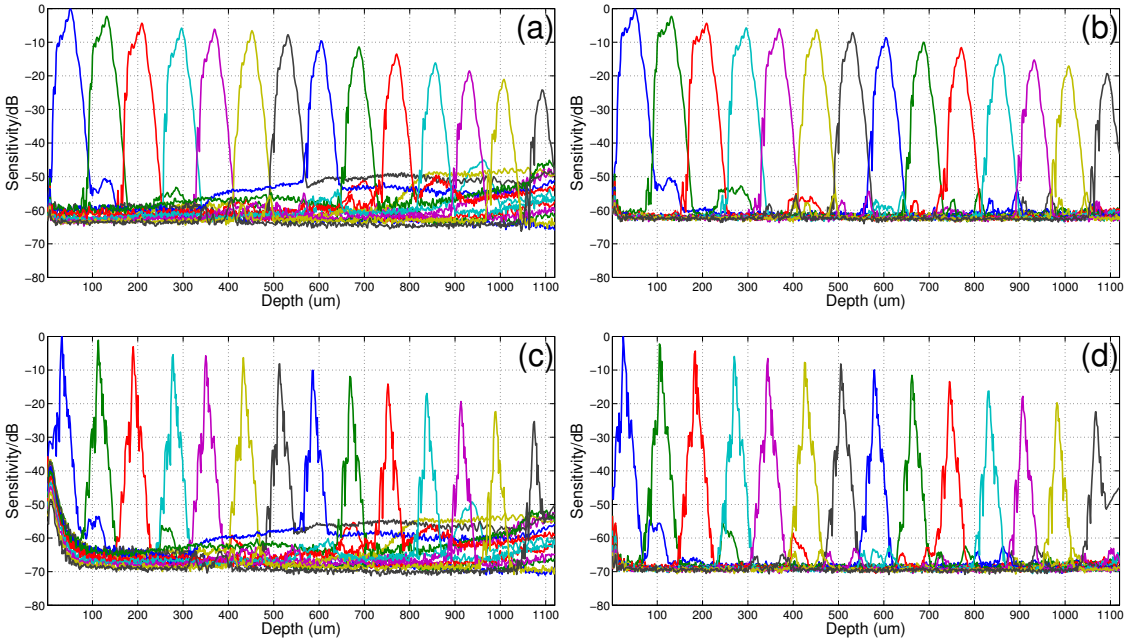

Figure 4.3: Sensitivity roll-off of systems applying (a) inverse UDFT to the linear wavenumber real spectra without dispersion compensation, (b) inverse NUDFT to the non-linear wavenumber real spectra without dispersion compensation, (c) dispersion compensation method in [1], (d) proposed dispersion compensation method on the non-linear wavenumber real spectra. A 2cm water cell is inserted to introduce large dispersion mismatch.

The above two calculations show that the influence of undesired term T_2 is relatively small compared with T_1 on the displayed half of reconstructed A-scan and that multiplying the correcting term directly to the non-linear wavenumber real spectra could achieve a satisfying dispersion compensation effect.

4.4.3 Incorporation of dispersion compensation to CS-FDOCT

Dispersion compensation method proposed in Sec. 4.4.1 can be written in matrix form as:

$$\begin{bmatrix} x_0^c \\ x_1^c \\ \vdots \\ x_{N-1}^c \end{bmatrix} = \begin{bmatrix} h(0,0) & h(1,0) & \dots & h(N-1,0) \\ \vdots & \vdots & \ddots & \vdots \\ h(0,N/2) & h(1,N/2) & \dots & h(N-1,N/2) \\ \vdots & \vdots & \ddots & \vdots \\ h(0,N-1) & h(1,N-1) & \dots & h(N-1,N-1) \end{bmatrix} \left(\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} .* \begin{bmatrix} e^{i\Phi(\omega_0)} \\ e^{i\Phi(\omega_1)} \\ \vdots \\ e^{i\Phi(\omega_{N-1})} \end{bmatrix} \right) \tag{4.10}$$

where $\mathbf{x^c} = [x_0^c, x_1^c, \dots, x_{N-1}^c]^T$ is the dispersion-compensated A-scan; ".*" stands for the component-wise multiplication. The transformation matrix is the inverse NUDFT matrix.

Denote $\exp(i\Phi(\omega_n))$ as Φ_n for $n \in [0, 1, \dots, N-1]$, then Eq. (4.10) can be rewritten by incorporating the dispersion compensation term into the inverse NUDFT matrix:

$$\begin{bmatrix} x_0^c \\ x_1^c \\ \vdots \\ x_{N-1}^c \end{bmatrix} = \begin{bmatrix} h(0,0)\Phi_0 & h(1,0)\Phi_1 & \dots & h(N-1,0)\Phi_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ h(0,N/2)\Phi_0 & h(1,N/2)\Phi_1 & \dots & h(N-1,N/2)\Phi_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ h(0,N-1)\Phi_0 & h(1,N-1)\Phi_1 & \dots & h(N-1,N-1)\Phi_{N-1} \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} \tag{4.11}$$

where $h(p,q)\Phi_n = \exp(i\omega_p \times q) \times \exp(i\Phi(\omega_n))$. Although the transformation matrix in Eq.

(4.11) does not have the same form as the inverse NUDFT matrix in Eq. (4.2), it could still be considered an inverse NUDFT matrix because the phases in each row are non-linear in wavenumber.

Equation (4.11) gives a transformation between the dispersion-compensated A-scan and the non-linear wavenumber real spectra, building the foundation of CS reconstruction. However, it cannot be used directly in CS-FDOCT because of decreased A-scan sparsity problem mentioned in Sec. 4.3.1. Thus the transformation matrix in Eq. (4.11) also needs modification to make $\mathbf{x}^c$ symmetric for higher sparsity:

$$\begin{bmatrix} x_0'^c \\ x_1'^c \\ \vdots \\ x_{N-1}'^c \end{bmatrix} = \begin{bmatrix} h(0,0)\Phi_0 & h(1,0)\Phi_1 & \dots & h(N-1,0)\Phi_{N-1} \\ h(0,1)\Phi_0 & h(1,1)\Phi_1 & \dots & h(N-1,1)\Phi_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ h(0,N/2-1)\Phi_0 & h(1,N/2-1)\Phi_1 & \dots & h(N-1,N/2-1)\Phi_{N-1} \\ h(0,N/2)\Phi_0 & h(1,N/2)\Phi_1 & \dots & h(N-1,N/2)\Phi_{N-1} \\ (h(0,N/2-1)\Phi_0)^* & (h(1,N/2-1)\Phi_1)^* & \dots & (h(N-1,N/2-1)\Phi_{N-1})^* \\ \vdots & \vdots & \ddots & \vdots \\ (h(0,1)\Phi_0)^* & (h(1,1)\Phi_1)^* & \dots & (h(N-1,1)\Phi_{N-1})^* \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} \tag{4.12}$$

The modification also preserves the value of desired half of the A-scan: $[x_0^c, x_1^c, \dots, x_{N/2-1}^c]^T = [x_0'^c, x_1'^c, \dots, x_{N/2-1}'^c]^T$.

The sensing matrix required for CS reconstruction can be obtained in the same way as in Section 3.2: take the conjugate transpose of the transformation matrix in Eq. (4.12):

$$\begin{bmatrix} (h(0,0)\Phi_0)^* & \dots & (h(0,N/2-1)\Phi_0)^* & \dots & h(0,N/2-1)\Phi_0 & \dots & h(0,1)\Phi_0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (h(N-1,0)\Phi_{N-1})^* & \dots & (h(N-1,N/2-1)\Phi_{N-1})^* & \dots & h(N-1,N/2-1)\Phi_{N-1} & \dots & h(N-1,1)\Phi_{N-1} \end{bmatrix} \tag{4.13}$$

With the under-sampled matrix of Eq. (4.13), dispersion compensated A-scan can be reconstructed from under-sampled non-linear wavenumber spectral data. It is noteworthy to mention that with proposed sensing matrix, dispersion compensation becomes a by-product of CS reconstruction and no additional dispersion compensation procedure is needed.

4.5 Experimental results

To evaluate the effect of proposed method, k-space data from a SDOCT system are used. The system uses a spectrometer having a 12-bit CMOS line scan camera (EM4, e2v, USA) with 2048 pixels at 70 kHz line rate. A superluminescent laser diode (SLED) is used as the light source which provides an output power of 10 mW and an effective bandwidth of $105nm$ centered at $845nm$. The experimental axial resolution of the system is $4.0\mu m$ in air while the transversal resolution is approximately $12\mu m$. All animal studies were conducted in accordance with the Johns Hopkins University Animal Care and Use Committee Guidelines. Spectral data were post-processed with MATLAB® R2012b on a desktop with Intel® Core^{TM} 2 Duo CPU (E8400, 3.0GHz), 4GB RAM, Windows® 7 64-bit operation system. The CS reconstruction algorithm is SPGL1 with default parameters [89, 90].

Proposed sensing matrix in Eq. (4.13) (denoted as the MNUDFT matrix) is applied to CS reconstruction on the under-sampled non-linear wavenumber real spectra. For comparison purpose, the following results are also evaluated: 1) original image obtained by applying NUDFT to 100% of the non-linear wavenumber real spectra; 2) image reconstructed using CS on the under-sampled non-linear wavenumber real spectra with the NUDFT matrix as

the sensing matrix.

The under-sampled non-linear wavenumber spectral data is obtained by applying a pseudo-random mask to the original spectra. Variable density random sampling [88] is used to generate this mask. This evaluation method is widely used in the studies of CS-SDOCT [30–32]. A CCD camera with randomly addressable pixels [30, 91–93] can be used to practically implement random under-sampling of spectral data in an SDOCT system.

The same under-sampled data, reconstruction domain and ε are used in the CS reconstructions of the same object. According to experimental results, OCT signal is sparse in spatial domain in most cases. However, it is usually sparser in wavelet domain than spatial domain. The reconstruction domain is chosen to optimize the reconstruction result. Besides, the sampling rate is chosen to balance the size of spectral data and the image quality while ε is selected to balance the loss of useful information and the reduction of noise. The dispersion compensation parameters a_2 and a_3 are set empirically.

To carry out a quantitative assessment of reconstruction results of different methods, the local contrast and signal to noise ratio (SNR) are computed. Their definitions are as follows:

$$\text{local contrast} = \frac{\mu_o}{\mu_b} \tag{4.14}$$

$$\text{SNR} = 20 \times \log_{10}\left(\frac{\sqrt{\frac{1}{N_o}\sum_{(i,j)\in\text{object}} I(i,j)^2}}{\sqrt{\frac{1}{N_b}\sum_{(i,j)\in\text{background}} I(i,j)^2}}\right) \tag{4.15}$$

where N_o and N_b are the number of pixels in the selected object region and background region respectively. As is shown in Figs. 4.4, 4.5, and 4.6, area in the red rectangles are

selected object region while the green rectangle area is selected background region. They have the same size. $I(i,j)$ is the intensity. μ_o and μ_b are mean of intensity of the object region and background region respectively.

The comparison is first implemented on the mouse paw scanning which contains several layers. The CS reconstructions use 40% spectral data; $\mathbf{W}$ is the four-level daubechies4 wavelet transform matrix. Compensation parameters are set as $a_2 = 157fs^2$ and $a_3 = 170fs^3$. All images are shown in the same dynamic range. Figure 4.4(b) exhibits bad quality because of the decreased A-scan sparsity problem. It achieves accurate reconstruction for pixels with high intensity but loses information for the pixels with relative low intensity. The layer beneath the surface is difficult to see due to the CS reconstruction error, as is pointed out by the arrows. Figure 4.4(c) has much better quality, which is very close to Fig. 4.4(a) which uses 100% sampling rate. Besides, Fig. 4.4(c) shows obvious dispersion compensation effect with clear and thin tissue boundary compared to Figs. 4.4(a) and 4.4(b). The overall contrast of Fig. 4.4(c) is better than that of Fig. 4.4(a) with 100% sampling rate because CS is well known to be good at reducing noise [23, 87]. The local contrast and SNR of Figs. 4.4(a), 4.4(b) and 4.4(c) are listed in Table 4.1 which shows that CS reconstruction with the MNUDFT matrix obtains better image quality.

Table 4.1: Local contrast and SNR of the B-scans of mouse paw in Fig. 4.4

	original	CS+NUDFT	CS+MNUDFT
Local contrast(dB)	4.63	3.50	**5.42**
SNR(dB)	14.06	11.44	**16.07**

The mouse cornea images are shown in Fig. 4.5. CS reconstructions used 37.5% sampling rate. $\mathbf{W}$ is the identity matrix. $a_2 = 120fs^2$ and $a_3 = 100fs^3$. Figure 4.5(c) is very close

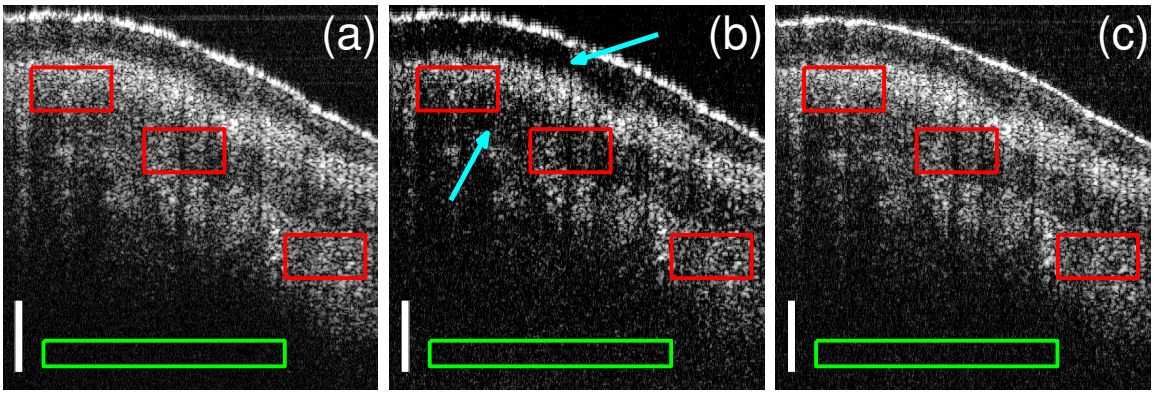

Figure 4.4: B-scans of a mouse paw. (a) original image obtained by applying NUDFT to 100% of the acquired non-linear wavenumber spectra; (b) CS reconstruction result with the NUDFT matrix from 40% of the acquired non-linear wavenumber spectra; (c) CS reconstruction result with the MNUDFT matrix from 40% of the acquired non-linear wavenumber spectra; The scale bars represent $100\mu m$. Image size in pixel is 450×1000.

to Fig. 4.5(a) while Fig. 4.5(b) shows obvious artifact and information loss. Regions of interest (ROI) are extracted from the reconstructed images (cyan rectangles in Figs. 4.5(a), 4.5(b) and 4.5(c)). Figure 4.5(f) shows that CS with the MNUDFT matrix preserves almost all the structures in the original image which uses 100% sampling rate while Fig. 4.5(e) are void of fine details.

Table 4.2 lists the local contrast and SNR of Figs. 4.5(a), 4.5(b) and 4.5(c).

Table 4.2: Local contrast and SNR of the B-scans of mouse cornea in Fig. 4.5

	original	CS+NUDFT	CS+MNUDFT
Local contrast(dB)	2.68	2.12	**3.50**
SNR(dB)	8.81	8.04	**9.67**

To give an in-depth assessment of the dispersion compensation effect of CS with the MNUDFT matrix, a 2.4cm water cell was placed in the reference arm when imaging a polymer-layered phantom to intentionally introduce a large dispersion mismatch between

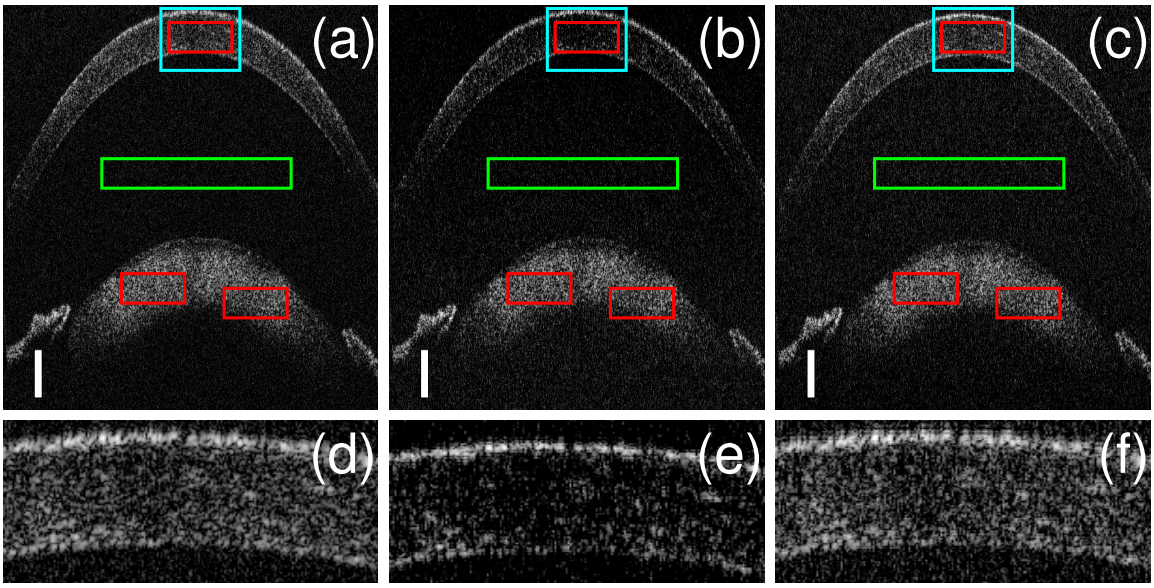

Figure 4.5: B-scans of a mouse cornea; (a) original image obtained by applying NUDFT to 100% of the acquired non-linear wavenumber spectra; (b) CS reconstruction result with the NUDFT matrix from 37.5% of the acquired non-linear wavenumber spectra; (c) CS reconstruction result with the MNUDFT matrix from 37.5% of the acquired non-linear wavenumber spectra; (d), (e) and (f) are zoom in of the cyan rectangle areas in (a), (b) and (c) respectively. The scale bars represent $100\mu m$. Image size in pixel is 700×1000.

the two arms. CS reconstructions use 50% sampling rate. $\mathbf{W}$ is the four-level daubechies4 wavelet transform matrix. $a_2 = 575fs^2$ and $a_3 = 295fs^3$. The B-scan obtained by applying the forward MNUDFT matrix to 100% of the acquired spectra (Eq. (4.12)) is added in Fig. 4.6(d) to show that MNUDFT can also achieve obvious dispersion compensation when used in the traditional imaging generation. Figure 4.6(a) shows much bigger dispersion than the previous cases. Figure 4.6(b) is void of fine details and there are obvious artifacts in the area outside the phantom. Dispersion compensated images, Figs. 4.6(c) and 4.6(d), are highly clear compared to Fig. 4.6(a), especially near the upper surface. Both of them show obvious dispersion reduction which validates the proposed method while CS uses only 50% of the acquired spectra and achieves better image quality, as is shown in Table 4.3.

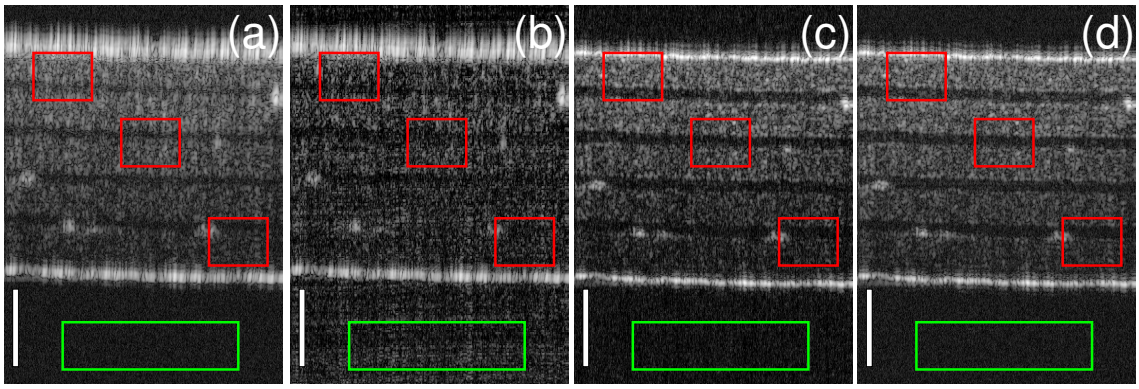

Figure 4.6: B-scans of a polymer-layered phantom with 2.4cm water induced dispersion; (a) original image obtained by applying NUDFT to 100% of the acquired non-linear wavenumber spectra; (b) CS reconstruction result with the NUDFT matrix from 50% of the acquired non-linear wavenumber spectra; (c) CS reconstruction result with the MNUDFT matrix from 50% of the acquired non-linear wavenumber spectra; (d) image obtained by applying the forward MNUDFT matrix to 100% of the acquired non-linear wavenumber spectra. The scale bars represent $100\mu m$. Image size in pixel is 450×1000.

Table 4.3: Local contrast and SNR of the B-scans of polymer-layered phantom in Fig. 4.6

	original	CS+NUDFT	CS+MNUDFT	MNUDFT
Local contrast(dB)	4.43	2.12	**7.71**	6.13
SNR(dB)	15.99	9.79	**27.23**	23.82

4.6 Discussion

The MNUDFT matrix is generated by "mirroring" its first half: setting columns of the right half to be the conjugate of the corresponding columns at the left half. This modification itself is not unique. Looking at the MNUDFT matrix in Eq. (4.13), one can easily create another sensing matrix by changing the order of the columns of its right half. Although the resulting A-scan will not be symmetric any more, this does not change the performance of the method since its $l1$-norm is unchanged and the displayed half of the

A-scan is preserved.

According to Eqs. (4.7) and (4.12), setting the entire bottom half of the transformation matrix to zero will maximize the sparsity of the undisplayed half of the A-scan. However, one cannot simply drop the information in this manner since it will make all the right half of the MNUDFT matrix zero. Denote the proposed MNUDFT matrix as $\mathbf{H}$. It implies $\mathbf{y} = \mathbf{H} * \tilde{\mathbf{x}}$ where $\mathbf{y}$ is the acquired spectra with non-linear wavenumber and $\tilde{\mathbf{x}}$ is the A-scan. If the right half of $\mathbf{H}$ are all zero, no matter what value the bottom half of $\tilde{\mathbf{x}}$ is (including the desired zero value), $\mathbf{y} = \mathbf{H} * \tilde{\mathbf{x}}$ does not hold any more if the first half of $\tilde{\mathbf{x}}$ is unchanged. Thus the basic principle is violated. In addition, the new sensing matrix should not have any two columns/rows the same, as is required by the standard CS theory [21, 22].

The proposition of the MNUDFT matrix starts from the forward NUDFT matrix instead of the strict inverse NUDFT matrix. This substitution does not change the effect of proposed method because the forward NUDFT matrix is only used to demonstrate what the A-scan would be after the modification. The usage of the forward NUDFT matrix to compute the A-scan has already been validated by the experiments in [79,82]. Then the MNUDFT matrix can be easily obtained because the modification is done on its conjugate transpose matrix. After all, the matrix that will be used in CS reconstruction as the sensing matrix is the MNUDFT matrix. Selected CS reconstruction algorithm relies on the sensing matrix and its conjugate transpose, not its inverse. Starting from the forward NUDFT matrix instead of the strict inverse NUDFT matrix is because it helps to demonstrate the motivation of proposed method.

Dispersion compensation coefficients a_2 and a_3 used in this paper were obtained empir-

ically from the system in the study. However, these two parameters can be obtained automatically using an iterative procedure which optimizes the sharpness of the reconstructed image, as in [1].

Another interesting subject is incorporating dispersion compensation directly into UDFT-based CS-FDOCT on under-sampled linear wavenumber spectral data. No interpolation is needed in this case. The Hilbert transform cannot be applied to under-sampled data. This can be overcome by multiplying the dispersion compensation term directly to the real spectra. But the remaining processing actually becomes CS reconstruction with the NUDFT matrix: the transformation equation on 100% spectral data ($\hat{\mathbf{y}}$) in this case is very similar to Eq. (4.11) except that the non-linear angular frequency ω is replaced by the linear angular frequency $\hat{\omega}$. Denote the transformation matrix as $\mathbf{D}$. Because $\Phi(\hat{\omega})$ is non-equispaced, the phase of arbitrary row in $\mathbf{D}$ is no longer equispaced. Thus $\mathbf{D}$ is a NUDFT matrix, which may introduce low sparsity to the resulting A-scan. Thus, modification on $\mathbf{D}$ is still needed to improve the sparsity of the resulting A-scan, especially when the dispersion is big.

We have shown that CS reconstruction with the proposed MNUDFT matrix on under-sampled non-linear wavenumber data represents a simpler approach compared to the traditional CS-FDOCT based on the UDFT matrix and under-sampled linear wavenumber data. The computation time required for the CS reconstructions in both cases are almost the same. However, other properties such as the difference in sampling rate and robustness to noise between UDFT and MNUDFT based CS-FDOCT still need more investigation.

There are some differences in the result when different CS reconstruction algorithms are used because the MNUDFT matrix is not unitary. Several CS reconstruction algorithms,

such as NESTA [94] and CSALSA [95], either cannot or are difficult to use to solve the CS optimization problem defined in Eq. (4.3), because they require either $\mathbf{F_u^T F_u} = \mathbf{I}$ or explicit $(\mathbf{I} - \mathbf{F_u^T F_u})^{-1}$ ($\mathbf{F_u^T}$ is the conjugate transpose of $\mathbf{F_u}$). In this work, SPGL1 is chosen as the reconstruction algorithm, in part because it does not require unitary sensing matrix.

Chapter 5

GPU-accelerated Realtime CS SD OCT

5.1 Realtime CS SD OCT on linear wavenumber sampling

5.1.1 Introduction

Reduced amount of k-space data required by CS can facilitate minimal data acquisition, increase imaging speed and decrease storage and transfer bandwidth needs. However, regardless of which CS reconstruction algorithm is used, the reconstruction of CS OCT image can take significantly more time compared to the reconstruction

of regular OCT image, for example, obtained at more than 80K A-lines per second [19, 51]. This has been the major hindrance to apply this imaging technique to clinical applications that typically require either real-time or immediate image reconstruction. In this section, we propose a practical method for achieving real-time CS SD-OCT imaging.

To achieve the goal of real-time CS SD-OCT imaging, we adopted parallel processing approach with graphics processing units (GPU) which has long been recognized as an effective way to accelerate computationally intensive task. CS reconstruction requires numerous matrix-vector multiplications which can be solved more efficiently by GPU than by CPU. GPU has been adapted to accelerate the CS reconstruction of various signals [53, 55, 57]. Compared to the CPU implementation, several orders of magnitude enhancement in speed has been commonly reported for the GPU based CS reconstruction.

In this section, we implemented real-time CS reconstruction of the spectral domain OCT (SD OCT) images using a triple-GPUs architecture. The CS reconstruction algorithm SpaRSA [96] is programmed through the NVIDIA's Compute Unified Device Architecture (CUDA) technology [60]. High quality SD OCT images can be reconstructed at > 70 frame/s, with the frame size 2048 (axial) $\times$ 1000 (lateral) and stopping iteration number 10. Compared to C++ and MATLAB implementations based on CPU, CS reconstruction using the triple-GPUs architecture achieved speed enhancements of 112 and 459 times respectively. The term "real-time" means that

the reconstruction of CS SD OCT is faster than data acquisition rate of the system. By using a CCD camera with randomly addressable pixels [91, 92], our program can achieve real-time reconstruction.

In CS OCT imaging, the A-scan image, $\mathbf{x}$ is obtained with high accuracy from under-sampled linear-in-wavenumber spectral data, $\mathbf{y}_u$ by solving the following unconstrained nonlinear convex optimization problem:

$$\underset{\mathbf{x}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{F_u x} - \mathbf{y_u}\|_\mathbf{2}^\mathbf{2} + \tau\|\mathbf{Wx}\|_\mathbf{1} \tag{5.1}$$

where $\mathbf{W}$ is the sparsifying operator. $\mathbf{F}_u$ is the under-sampled Fourier transform matrix. τ is the regularization parameter that controls the sparsity of the reconstructed A-scan. The notation $\|\mathbf{c}\|_1$ is the l_1 norm, which is defined as $\|\mathbf{c}\|_1 = \sum_i |c_i|$. $\| \bullet \|_2$ is the l_2 norm. In this section, $\mathbf{W}$ is chosen to be the identity matrix because OCT signals are usually sparse enough in spatial domain [30, 31].

5.1.2 Method

The selected CS reconstruction algorithm is SpaRSA [96] which tries to solve Eq.(5.1) through an iterative procedure. In each iteration, SpaRSA obtains the new iterate $\mathbf{x}^{k+1}$ from the current iterate $\mathbf{x}^k$ by solving the following sub problem:

$$\mathbf{x}^{k+1} = \min_{\mathbf{z}}\ \frac{1}{2}\|\mathbf{z} - \mathbf{u}^k\|_2^2 + \frac{\tau}{\alpha_k}\|z\|_1 \tag{5.2}$$

for some $\alpha_k > 0$. $\mathbf{u}^k = \mathbf{x}^k - (1/\alpha_k)\mathbf{F}_u^T(\mathbf{F}_u\mathbf{x}^k - \mathbf{y}_u)$. $\mathbf{F}_u^T$ is the adjoint matrix of $\mathbf{F}_u$. Equation (5.2) can be viewed as the quadratic separable approximation of Eq.(5.1) (up to a constant) [96]. Equation (5.2) can be solved separately in the component of $\mathbf{z}$:

$$x_i^{k+1} = \min_{z_i} \frac{1}{2}(z_i - u_i^k)^2 + \frac{\tau}{\alpha_k}|z_i| = \text{soft}(u_i^k, \frac{\tau}{\alpha_k}) \tag{5.3}$$

for $i \in [0, 1, \ldots, N-1]$. N is the length of $\mathbf{x}$. soft(x,a)$\hat{=}$max$\{|x|-a, 0\} \times x/$max$\{|x|-a, 0\}$ is the complex soft-threshold function.

Inspired by Barzilai and Borwein [97], α_k is chosen as the approximation of the Hessian $\mathbf{F}_u^T\mathbf{F}_u$ in [96]:

$$\alpha_k = \|\mathbf{F}_u\mathbf{s}^k\|_2^2/\|\mathbf{s}^k\|_2^2 \tag{5.4}$$

where $\mathbf{s}^k = \mathbf{x}^k - \mathbf{x}^{k-1}$.

The computation procedure for the reconstruction of an A-scan on a GPU is shown in Fig. 5.1. β is the desired stopping iteration number. k is the current iteration number. $\mathbf{I}$ is the sampling mask corresponding to $\mathbf{y}_u$. The procedure in Fig. 5.1 is different from [96] in that it uses a different stopping criterion. The reconstruction stops when the desired iteration number is achieved to avoid data divergence, which will be discussed later.

Each iteration includes the computationally extensive tasks such as one matrix-vector multiplication each by $\mathbf{F}_u^T$ and $\mathbf{F}_u$ for $\mathbf{u}^k$, N operations to solve Eq.(5.3), and two l_2-norm computation for α_k ($\mathbf{F}_u\mathbf{s}^k$ can be obtained from the intermediate values:

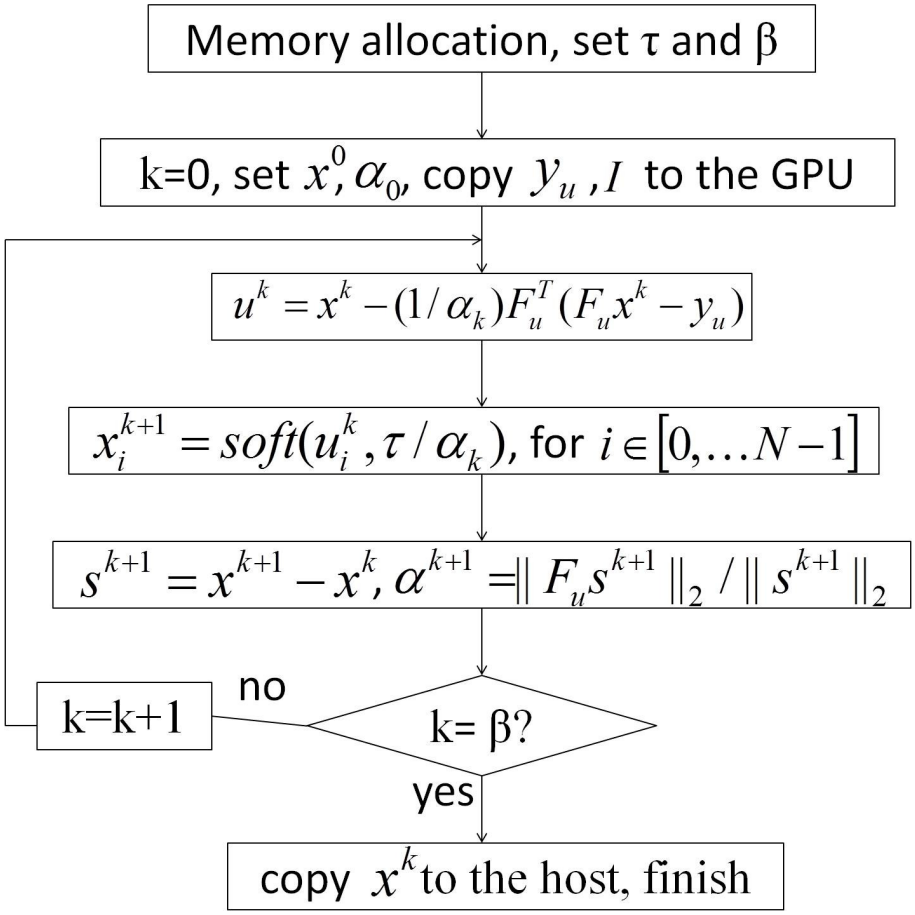

Figure 5.1: reconstruction procedure of an A-scan.

$\mathbf{F}_u\mathbf{s}^k = \mathbf{F}_u\mathbf{x}^k - \mathbf{F}_u\mathbf{x}^{k-1}$). All of these can be solved efficiently through GPU.

Since the reconstructions of CS OCT A-scans are independent of each other, to maximize the computational power of a single GPU and smooth the B-scan display, our approach reconstructs the A-scans in one B-scan (1000 lines) together instead of sequentially.

Every computation step in Fig.5.1 is applied to all A-scan data simultaneously. E.g. for each iteration, our program computes the α_k for 1000 A-scans simultaneously. Thus this approach maintains a vector of α whose length is the number of A-scans.

This is different from the case in which one CUDA thread reconstructs one A-scan, which limits the thread number and does too many computations on one CUDA core. The under-sampled raw A-scan spectral data are obtained using a common sampling mask.

The matrix-vector multiplications of $\mathbf{F}_u$ and $\mathbf{F}_u^T$ and its efficiency is critical to the speed of the reconstruction. Although the matrix-matrix multiplication operator in the CUBLAS library achieves significant acceleration, it is still too slow for achieving real-time imaging. Inspired by [57], our program takes advantage of the CUFFT library. For every A-scan, $\mathbf{F}_u\mathbf{x}^k$ is computed in two steps: (1) $\mathbf{t}^k = FFT(\mathbf{x}^k)$; (2) under sample $\mathbf{t}^k$ with $\mathbf{I}$. $\mathbf{F}_u^T(\mathbf{F}_u\mathbf{x}^k - \mathbf{y}_u)$ is computed in a similar way: (1) $\mathbf{r}^k = \mathbf{F}_u\mathbf{x}^k - \mathbf{y}_u$; (2) zero-padding $\mathbf{r}^k$ according to $\mathbf{I}$, denote the result as $\mathbf{t}^k$; (3) compute the $IFFT(\mathbf{t}^k)$. $\mathbf{t}^k$ has the same length as $\mathbf{x}$ in both cases. This method can be easily adapted to the multiplication of $\mathbf{F}_u$ and $\mathbf{F}_u^T$ to the data of multiple A-scans since CUFFT provides the FFT operator for batch execution of multiple one-dimensional transform. Experimental results show that our implementation of the matrix-matrix multiplication is more than 10 times faster than the CUBLAS version (mainly due to the speed advantage of FFT). For different sampling size (size of $\mathbf{y}_u$), the $FFT/IFFT$ operator is applied to the same size data; thus the sampling rate has little effect on the reconstruction speed. The proper sampling rate should be chosen to balance the image quality and input data size. Another benefit of our implementation of the matrix-matrix multiplication is that instead of the whole sensing matrix, only the

one-dimensional sampling mask ($\mathbf{I}$) is stored in the GPU which is more compact.

l_2-norm operator in the CUBLAS library cannot compute the l_2-norm of multiple A-scan vectors simultaneously. We wrote a kernel which applies the tree-based reduction in shared memory, as advocated in reference [98]. Our implementation computes the l_2-norms in Eq. (5.4) of all the A-scans simultaneously and is more than 100 times faster than computing the l_2-norm of the A-scans sequentially using the CUBLAS l_2-norm operator.

The one-dimensional complex soft-threshold function as well as the vector additions/subtractions can be easily implemented in parallel for multiple A-scan reconstructions on the GPU.

The stopping criterion in our program is different from that in [96]. The reconstruction of a B-scan stops when the desired iteration number (β) is achieved. In this way, the reconstruction time for any B-scan is determined solely by β and is independent of the input data as well as other parameters. When reconstructing a sequence of B-scans, fixed β will smooth the reconstruction rate. Besides, when reconstructing multiple A-scans simultaneously, inconsistent iteration numbers among A-scans are difficult to control and create a large number of divergence branches that slows the GPU programs a lot.

Although the reconstructions of the A-scans in a B-scan stop together at β iteration, some A-scans may converge in less than β iteration while some requires more. The former case does not influence the reconstruction result or the reconstruction

time. The latter case will degrade the image quality and can either be solved by increasing β or attenuated by using a sequence of decreasing τ for every iteration [57,96]. Choosing β is a trade-off between the reconstruction speed and image quality. According to our experience, setting β to 10-20 is usually enough for a high-quality reconstruction; however, how to optimize β is still under study.

The regulation parameter τ determines the sparsity of the result as well as the image quality. The performance of SpaRSA generally degrades for smaller τ [57, 96], while larger τ will lead to better noise reduction and more loss of low-contrast features. As is stated above, better choice of τ will accelerate the reconstruction process by making the reconstruction converge in less iteration. How to optimize τ is still an open field in the CS reconstruction of real signals.

To achieve larger acceleration, we implemented CS reconstruction using three GPUs. The signal processing flow chart of the triple-GPUs architecture is illustrated in Fig. 5.2. Four major threads exist: three for GPU control and one for display.

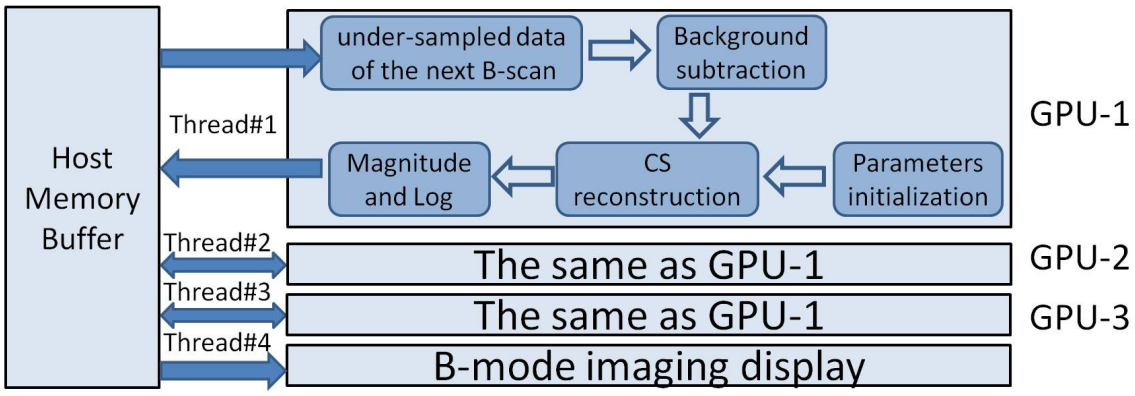

Figure 5.2: Triple-GPUs architecture.

Each GPU is controlled by one thread which reads the under-sampled data of the next B-scan, subtracts the DC term, performs CS reconstruction, then copy the result to the host computer. The fourth thread displays the B-scan results in the order of the frame number. GPUs with similar computational power are desired to smooth the display.

5.1.3 Results and Discussion

K-space data from an SD OCT system is used to evaluate the program. The system uses a spectrometer having a 12-bit CMOS line scan camera (EM4, e2v, USA) with 2048 pixels at 70kHz line rate. The light source is a superluminescent laser diode (SLED) with output power of 10mW and effective bandwidth of $105nm$ centered at $845nm$. The experimental axial resolution of the system is $4.0\mu m$ in air while the transversal resolution is approximately $12\mu m$. All animal studies were conducted in accordance with the Johns Hopkins University Animal Care and Use Committee Guidelines.

A workstation with Intel Core i7 CPU (3.6GHz), 16GB RAM, Windows7 64-bit operation system is used as the host computer which contains three GPUs: GPU-1 (NVIDIA GeForce GTX 670) with 1344 stream processors, 1GHz processor clock and 2GB global memory; GPU-2 (NVIDIA Tesla C2075) with 2496 stream processors, 1.15GHz processor clock and 5GB global memory; GPU-3 which is the same as GPU-2.

For comparison, our program is also implemented on the single-GPU architecture (GPU-1) and dual-GPUs architecture (GPU-2 and GPU-3). The implementation on the single GPU architecture is similar to that on the triple-GPUs architecture except only one thread for GPU control exists. The dual-GPUs case has two threads for GPU control. C++ and MATLAB (version R2013a) implementations on CPU are also compared. They reconstruct the A-scans sequentially. The speed of our programs is evaluated with the benchmark line rate test. The same under-sampled data, parameters and algorithm (Fig. 5.1) are used in the comparison and the same B-scan results are obtained in all implementations with different speed. The FFT operator from $FFTW$ [99] is used in the C++ implementation while MATLAB uses its built-in FFT operator. C++ program is not fully optimized (only compiled with the flag '-O2'). The speed of CS reconstruction with the CPU-based multi-threading will be studied in future work. GPU programs are optimized with the CUDA Visual Profiler.

To show the computation ability of the proposed method, the under-sampled linear-in-wavenumber data set is obtained by sampling from the pre-generated k-space data set with a calibrated k-linear sampling mask [31] and stored in the RAM. This is consistent with the desired practical application in which the under-sampled data will be grabbed from the camera and stored in the RAM. The starting values are set as $\mathbf{x}^0 = \mathbf{0}$ and $\alpha_0 = 1$ for all A-scans. These two values do not show obvious influence on the image quality and computation time.

We first evaluate all the implementations on the reconstructions of OCT images of an orange. The sampling rate is 40%. $\tau = 2.5$. The stopping iteration number is 10. The reconstruction result with 100% samples is provided in Fig. 5.3(a) as a reference. The CS reconstruction result is shown in Fig. 5.3(b). Both cases are averaged on 100 frames. As is shown by the images, the CS reconstruction result is very close to the reference image. For a quantitative assessment, the peak signal to noise ratio (PSNR) is computed for both images with definition $PSNR = 10\log_{10}(\max^2(f(\mathbf{x}))/var)$. var is the intensity variance of the selected background region (the white dash rectangle areas in Figs. 5.3(a) and 5.3(b)). $f(\mathbf{x})$ is the B-scan. The PSNR of Figs. 5.3(a) and 5.3(b) are 66.17dB and 75.32dB respectively. The CS reconstruction achieves 9.15dB better PSNR; it is well known to be good at reducing noise [30, 87].

The comparison of the line rate versus stopping iteration number for different implementations is shown in Fig. 5.3(c). The data were obtained by applying these implementations to the reconstruction of the orange image. $\tau = 2.5$ and sampling rate is 40%. The speedup of the implementation on the triple-GPUs architecture compared to the other implementations is illustrated in Fig. 5.3(d). Speedup is defined as T_o/T_t where T_t is the single B-scan reconstruction time with triple-GPUs and T_o is that of the other implementations. The triple-GPUs implementation achieves an average of 459.36, 112.02, 2.83, and 1.77 times speedup compared to those based on the MATLAB, C++, single-GPU and dual-GPUs respectively. The data in Figs. 5.3(c), 5.3(d) are averaged in 30 runs.

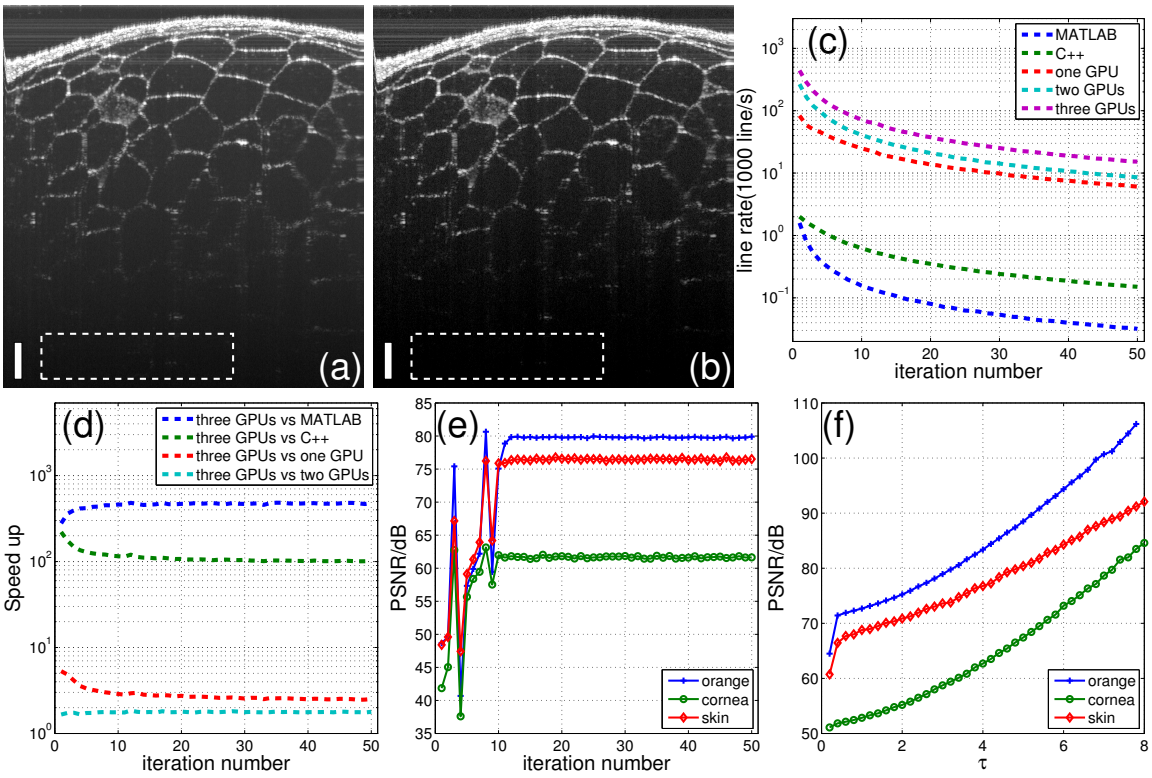

Figure 5.3: (a) original OCT image of an orange; (b) CS reconstruction result; (c) line rate vs. iteration number; (d) speedup vs. iteration number; (e-f) PSNR vs. iteration number and τ respectively for B-mode imaging of orange, human skin, and mouse cornea. The scale bars in (a) and (b) represent $100um$.

We test the program on the realtime processing of OCT images of orange, human skin, and mouse cornea. The image frames rates in these three cases are all 72 fps with stopping iteration number 10, $\tau = 3.5$ and sampling rate 45%. This shows that the speed of our program is independent of the imaging object which enables the development of a system that incorporates the CS technique.

To illustrate the influence of the iteration number on the image quality, the PSNR is computed for the reconstructed images of an orange, human skin and mouse cornea with different iteration numbers. τ and sampling rate are stated above. The back-

ground regions of the human skin and mouse cornea are selected similarly to that of the orange. PNSR vs iteration number for these three samples is displayed in Fig. 5.3(e). As it shows, the PSNR usually converges between 10 to 20 iterations; thus setting the stopping iteration number to 10-20 (corresponding to 38-72 fps) is usually enough for a high-quality image.

Fig. 5.3(f) shows the PSNR versus τ for all three samples. The stopping iteration number is 10 and the sampling rate is stated above. Higher τ results in a better denoising effect and more loss of low intensity features.

5.2 GPU-accelerated non-uniform fast Fourier transform-based CS SDOCT

5.2.1 Introduction

In Chapter 4, we show that the under-sampled nonlinear wavenumber sampling can be used directly in the CS SD OCT by taking the modified non-uniform discrete Fourier transform (MNUDFT) matrix as the sensing matrix, instead of using UDFT/FFT. However, CS with the MNUDFT matrix is extremely time-consuming if the matrix multiplication is used in the iterative reconstruction algorithm, since the computation complexity of the matrix multiplication is $O(N \times M)$ where M and N ($M < N$) are the row and column number of the under-sampled MNUDFT matrix,

respectively.

The gridding-based non-uniform fast Fourier transform (NUFFT) [100] has been shown to be a good approximation to NUDFT and offers better performance than the interpolation+FFT method [79,101–103]. NUFFT reduces the computation complexity to $\sim O(Nlog(N))$ and has already been adapted for the signal processing of regular SD OCT [79,101–103]. As shown in Section 4.3.1, the sparsity problem for the NUDFT matrix that excludes its usage in the CS SD OCT does not exist for CS with NUFFT. NUFFT is indeed a faster implementation that approximates MNUDFT when used in the CS SD OCT. By adding the zero-padding/under-sampling steps to the CS reconstruction procedure, NUFFT can be used in the CS SD OCT with the under-sampled nonlinear wavenumber sampling.

Compared to the reconstruction of regular OCT image, CS SD OCT takes significantly more time which hinders its applications in the surgical conditions. However, such computation is ideal for GPU acceleration which can significantly reduce the computation time. In this section, we implemented the GPU-accelerated NUFFT-based CS SD OCT on the same triple-GPUs architecture as Sec. 5.1 and demonstrated that real-time B-mode imaging can be reconstructed at more than 30 fps with frame size 2048(axial)×1000(lateral). Both Kaiser-Bessel (KB) and Gaussian function are tested independently as the convolution kernel for NUFFT with different over-sampling ratios and kernel widths. The computation speed is evaluated in both the pre-computed mode and the on-the-fly mode. We also implement the GPU-

accelerated MNUDFT matrix-based CS SD OCT as a comparison for speed and image quality. The experimental results show that GPU-NUFFT-CS is a good approximation to the GPU-MNUDFT-CS in terms of the image quality while providing more than 5 times speed enhancement. Compared to GPU-FFT-CS, GPU-NUFFT-CS has improved sensitivity roll-off, better signal-to-noise ratio, and less side-lobes while eliminating the need for the cumbersome k-space grid filling and the pre-calibration process.

5.2.1.1 NUFFT

The gridding-based NUFFT is a fast algorithm that approximates NUDFT [100] and has been shown to offer better performance than the interpolation+FFT method [79, 101–103]. Two different kinds of NUFFT will be discussed in this section. Following the definition in [100], they are called type-1 and type-2 NUFFT. The type-1 NUFFT transforms the nonlinear wavenumber data to a linear spatial representation while type-2 NUFFT transforms the linear spatial data to a nonlinear wavenumber representation. They are a mutually inverse transformation.

The computation steps for type-1 NUFFT are summarized in the following:

(1) Compute the new coefficients on an over-sampled uniform grid by convolution with a kernel function:

$$F_r(\hat{k}_i) = \sum_n F(k_n) G(\hat{k}_i - k_n), i = 0, \ldots, M_r - 1 \tag{5.5}$$

where $\hat{\mathbf{k}} = [\hat{k}_0, \ldots, \hat{k}_{M_r-1}]$ is the uniform grid covering the same range as $\mathbf{k}$. $F_r(\hat{k}_i)$ is the new coefficient with linear wavenumber while $F(k_n)$ is the nonlinear wavenumber data. $G(\kappa)$ is the kernel function. $M_r = R \times N$ is the size of linear wavenumber grid. R is the over-sampling ratio. Usually, Eq. (5.5) is evaluated within a window defined by the kernel width W for each $\hat{k}_i$: the range of n satisfies $|k_n - \hat{k}_i|/\delta\hat{k} \leq W/2$, $\delta\hat{k} = \Delta K/M_r$.

(2) Apply regular FFT to the new coefficients $F_r(\hat{\mathbf{k}})$.

(3) Deconvolve the resulting signal by a division of the Fourier transform of the kernel function:

$$\tilde{f}_r(x_m) = f_r(x_m)/g(x_m), m = 0, \ldots, M_r - 1 \tag{5.6}$$

where $f_r(\mathbf{x})$ and $g(\mathbf{x})$ are the Fourier transform of $F_r(\hat{\mathbf{k}})$ and $G(\kappa)$, respectively.

(4) Truncate the signal to the original size.

The computation steps for type-2 NUFFT are also summarized:

(1) Zero-pad the signal to the length of the over-sampling grid.

(2) Deconvolve the signal by a division of the Fourier transform of the kernel function (Eq. (5.6)).

(3) Apply regular FFT to the resulting signal.

(4) Compute the coefficients with nonlinear wavenumber using the kernel convolution:

$$F(k_n) = \sum_i F_r(\hat{k}_i) G(k_n - \hat{k}_i), n = 0, \ldots, N - 1 \tag{5.7}$$

The same as Eq. (5.5), Eq. (5.7) is usually evaluated within a window with the same width W for each k_n.

Both the Gaussian function [79, 102] and Kaiser-Bessel (KB) function [101, 103] have been used as the convolution kernel in the processing of the regular SD OCT signal with NUFFT. KB kernel is found to be a better choice in terms of image quality [82,101,103] while Gaussian kernel offers a better accuracy-speed trade-off [103]. The definition of the Gaussian kernel and its Fourier transform are [79, 82, 100, 102, 103]:

$$G(\kappa) = \exp(-\frac{\kappa^2}{4\theta}) \tag{5.8}$$

$$g(x_m) = \sqrt{2\theta}\exp(-x_m^2\theta) \tag{5.9}$$

The optimal value for θ is [100]:

$$\theta = \frac{1}{N^2}\frac{\pi}{R(R-0.5)}\frac{W}{2} \tag{5.10}$$

The definition of KB kernel and its Fourier transform are listed as follows [82,103, 104]:

$$G(\kappa) = I_0\left(\beta\sqrt{1-(2\kappa/W)^2}\right)/W \tag{5.11}$$

where $I_0(\kappa)$ is the zero-order modified Bessel function of the first kind.

$$g(x_m) = \frac{\sin\left(\sqrt{(m\pi W/M_r)^2-\beta^2}\right)}{\sqrt{(m\pi W/M_r)^2-\beta^2}} \tag{5.12}$$

The optimal value for β is given by [82, 104]:

$$\beta = \pi \sqrt{\frac{W^2}{R^2}\left(R - 0.5\right)^2 - 0.8} \tag{5.13}$$

The over-sampling ratio R and kernel width W influence not only the computation complexity or the speed, but also the accuracy of the reconstruction, regardless of the kernel used. Increasing R and W will reduce the truncation error introduced by using a finite width window [103]. It has been reported [82, 104] that R between 1 and 2 is allowed, which shows the potential to reduce the computation time while retaining the image quality. In this paper, NUFFT-CS is tested independently using both the Gaussian and KB function with different sampling ratios (R=1.5 and 2) and kernel widths (W=3 and 5).

5.2.1.2 NUFFT-CS

To incorporate NUFFT into CS reconstruction, two issues need to be resolved: the first one is that CS demands the under-sampled version of NUFFT. Both the input and output signals of type-1 and type-2 NUFFT have the length N. Thus, during the CS reconstruction, we cannot apply type-1 NUFFT to $\mathbf{y}_u$ for $\mathbf{F}_u^T\mathbf{y}_u$ or type-2 NUFFT to $\mathbf{x}$ for $\mathbf{F}_u\mathbf{x}$. $\mathbf{F}_u^T$ is the adjoint matrix of $\mathbf{F}_u$. Instead of computing the under-sampled versions of type-1 and type-2 NUFFT, we use a similar procedure as in Sec. 5.1 to solve this. Denote $\mathbf{I}$ as the sampling mask corresponding to $\mathbf{y}_u$.

$\mathbf{F}_u\mathbf{x}$ is computed in two steps: (1) apply the type -2 NUFFT to $\mathbf{x}$, denote the result as $\mathbf{t}$; (2) under-sample $\mathbf{t}$ with $\mathbf{I}$. $\mathbf{F}_u^T\mathbf{y}_u$ is computed in a similar way: (1) zero-padding $\mathbf{y}_u$ to length N according to $\mathbf{I}$, denote the result as $\mathbf{y}_z$; (2) apply the type-1 NUFFT to $\mathbf{y}_z$. Clearly, this implementation does not change the result of $\mathbf{F}_u\mathbf{x}$ and $\mathbf{F}_u^T\mathbf{y}_u$. In fact, this implementation can be considered as a convenient interface for incorporating any transformation into CS reconstruction. No under-sampled version of that transformation needs to be computed. It is especially useful in some cases like NUFFT: the under-sampled version of the transformation may not exist or its closed form is difficult to compute.

Another issue for NUFFT-CS is to check to see whether NUFFT has a similar sparsity problem as NUDFT. NUFFT is a good approximation to the NUDFT. If its resulting A-scan has a similar A-scan sparsity as the one from NUDFT, NUFFT cannot be applied to CS reconstruction. We compared the A-scans obtained by applying NUDFT, MNUDFT, type-1 NUFFT to a full-length nonlinear wavenumber data (belonging to a multi-layer polymer phantom scanning) and FFT to the full-length linear wavenumber data (obtained from the full-length nonlinear wavenumber data using cubic interpolation). The results are shown in Fig. 5.4.

The A-scan obtained from NUFFT has similar sparsity as the A-scans from MNUDFT and FFT. All of them are much higher than the sparsity of the A-scan using NUDFT. In SD OCT, only the left half of the A-scan is displayed. The displayed halves of Figs. 5.4(a)–5.4(c) are very close. However, for the right half, the

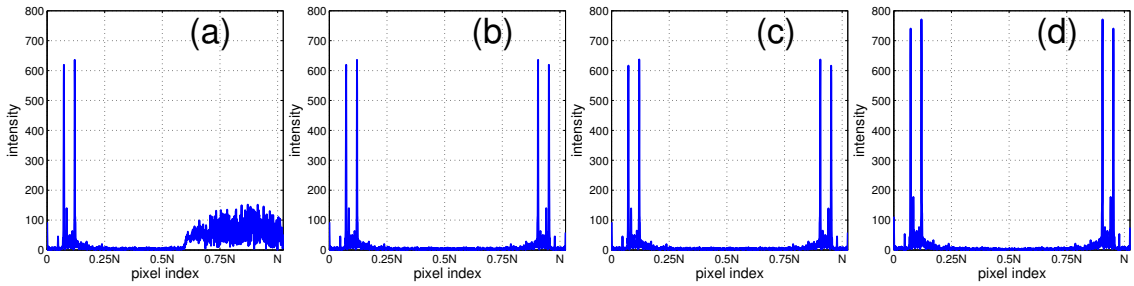

Figure 5.4: Sparsity comparison of A-scans by applying (a) NUDFT, (b) MNUDFT, (c) type-1 NUFFT to the full-length nonlinear wavenumber spectral data, and (d) FFT to the full-length linear wavenumber spectral data.

NUFFT result has much higher sparsity than NUDFT and is symmetric to the left half, which is the same as MNUDFT. The reason is that NUFFT is based on FFT. If the input data is in real number, after the computation step (1) and (2), the data is still real. Symmetric structure signal will be achieved when applying FFT to real number vector, as described in step (3). Step (4) simply removes some redundant data; thus, its resulting signal will be symmetric. There is no sparsity problem for NUFFT-CS. In fact, it is easy to figure out that NUFFT is a good approximation to MNUDFT in this application, not NUDFT. Although a special case is used in Fig. 5.4, the same phenomenon will happen when applying the above four transforms to other SD OCT data.

Thus by adding the zero-padding/under-sampling steps to the CS procedure, NUFFT can be used in CS SD OCT with the under-sampled nonlinear wavenumber sampling.

5.2.2 Method

Here, we describe the implementation of the GPU-accelerated NUFFT-based CS reconstruction, as well as the GPU-accelerated MNUDFT-based CS reconstruction for the purpose of comparing the image quality and speed. All the programs are implemented through NVIDIA's Compute Unified Device Architecture (CUDA) technology [60].

The implementation of CS algorithm SpaRSA on a GPU has been demonstrated in Sec. 5.1. The computation procedure of SpaRSA is demonstrated below. L is the number of A-scans in one B-scan.

The thread number in our implementation is 128. It is tuned using the NVIDIA's Visual Profiler for the highest reconstruction speed. The block number is chosen based on the size of the data. For example, for the operation in line 15 of the computation procedure, the block size will be the smallest integer bigger than $2048\times1000/128$. Here, 2048 is the size of the A-scan; 1000 is the number of A-scans in one B-scan. Thus, the block number changes for different operations in our implementation.

5.2.2.1 GPU-MNUDFT-CS

The implementation of $\mathbf{F}_u\mathbf{x}$ and $\mathbf{F}_u^T\mathbf{r}$ with NUFFT has been discussed above. The CUFFT library is used for the batched FFT transform. Another benefit of this implementation is that instead of the whole sensing matrix only the one-dimensional binary sampling mask $\mathbf{I}$ is stored in the GPU memory, which is more compact.

Computation procedure of SpaRSA for a B-scan on one GPU

Input:

$\mathbf{y}_u^1, \ldots, \mathbf{y}_u^L$ # under-sampled data for each A-scan
$\mathbf{I}$ or $\mathbf{F}_u$ # sampling mask or the sensing matrix
η # stopping iteration number
τ # CS regulation parameter

Output:

$(\mathbf{x}^1)^\eta, \ldots, (\mathbf{x}^L)^\eta$ # $(\mathbf{x}^i)^j$ is the j-th iterate of the i-th A-scan

Initialization: set $(\alpha^1)^0, \ldots, (\alpha^L)^0$, $(\mathbf{x}^1)^0, \ldots, (\mathbf{x}^L)^0$
copy $\mathbf{y}_u^1, \ldots, \mathbf{y}_u^L$, $\mathbf{I}$ or $\mathbf{F}_u$ to GPU
$p \leftarrow 0$
for all $j \in [1, \ldots, L]$ **do in parallel**
    $(\mathbf{r}^j)^0 = \mathbf{F}_u(\mathbf{x}^j)^0 - \mathbf{y}_u^j$
end for
while $p \neq \eta$ **do**
    for all $j \in [1, \ldots, L]$ **do in parallel**
        $(\mathbf{u}^j)^p = (\mathbf{x}^j)^p - (1/(\alpha^j)^p)\mathbf{F}_u^T(\mathbf{r}^j)^p$
    end for
    for all $i \in [0, \ldots, N-1]$ **and** $j \in [1, \ldots L]$ **do in parallel**
        $(x_i^j)^{p+1} = \text{soft}\big((u_i^j)^p, \tau/(\alpha^j)^p\big)$ # x_i^j is the i-th component of $\mathbf{x}^j$
    end for
    for all $j \in [1, \ldots, L]$ **do in parallel**
        $(\mathbf{s}^j)^{p+1} = (\mathbf{x}^j)^{p+1} - (\mathbf{x}^j)^p$
        $(\mathbf{r}^j)^{p+1} = \mathbf{F}_u(\mathbf{x}^j)^{p+1} - \mathbf{y}_u^j$
        $(\alpha^j)^{p+1} = \|(\mathbf{r}^j)^{p+1} - (\mathbf{r}^j)^p\|_2 / \|(\mathbf{s}^j)^{p+1}\|_2$
    end for
    $p \leftarrow p + 1$
end while
Finish: copy $(\mathbf{x}^1)^\eta, \ldots, (\mathbf{x}^L)^\eta$ to the host memory

For SD OCT, the nonlinear wavenumber $\mathbf{k}$ is static. The value of $G(\kappa)$ and $g(\mathbf{x})$ can be pre-computed. Also, the indices of the uniform grid (range of n) for each $\hat{k}_i$ in Eq. (5.5) and the indices of the nonlinear wavenumber (range of i) for each k_n in Eq. (5.7) can be obtained before-hand. The computation time then is independent of the choice of kernel function.

Although our program is tested on the reconstruction of SD OCT signal, it can also be used in the swept source OCT (SS OCT) that has nonlinear wavenumber. In the case that frequency jitters or low repeatability of frequency scanning happens in SS OCT, pre-computation of the NUFFT parameters are not feasible and will degrade the reconstruction accuracy if they are used in the reconstruction. Instead, the NUFFT parameters should be computed on-the-fly. In this work, the GPU-NUFFT-CS with the Gaussian kernel is implemented in both the pre-computed mode and on-the-fly mode while that with the KB kernel is only implemented in the pre-computed mode because the zero-order modified Bessel function of the first kind has not been incorporated to CUDA. The results obtained in the pre-computed mode and on-the-fly mode are the same for GPU-NUFFT-CS with the Gaussian kernel.

5.2.2.2 GPU-NUFFT-CS

The implementation of $\mathbf{F}_u\mathbf{x}$ and $\mathbf{F}_u^T\mathbf{r}$ with the MNUDFT matrix is straightforward. MNUDFT matrix should be under-sampled according to the sampling mask first before being transferred to the GPU. The matrix multiplication operator in CUBLAS

is used in the program.

5.2.3 Results and Discussion

Our methods are evaluated using the spectral data from an SD-OCT system. A superluminescent laser diode (SLED) with a center wavelength of 845 nm and a bandwidth of 105 nm is used as the light source. The spectrometer contains a 12-bit CMOS line scan camera (EM4, e2v, USA) with 2048 pixels at 70 kHz line rate. The beam scanning speed in our experiment is 70mm/s. The experimental axial resolution of the system is 4.0 μm in air and the transversal resolution is approximately 12 μm. The sensitivity of our system is 92.5dB, which is tested by placing a mirror as a sample at the position of 80μm from the zero-delay (corresponding to the first peaks in Figs. 5.5(a)-5.5(f)) and a ND filter (-35.5dB) in front of the mirror while the reference level was set at the level for shot-noise limited detection. It is computed using the method in [49]. For all experiments, the power of light incidents on the samples is 3.7mW, which is conformity with the American National Standard Institute (ANSI) standard for safe use of lasers [105] on the experimental samples in our work.

The experimental data in this section is measured on a workstation with an Intel Xeon CPU (3.46 GHz), 16 GB memory and a NVIDIA GeForce GTX 580 GPU with 512 stream processors, 1.5 GHz processor clock and 1.5 GB global memory. The under-sampled nonlinear wavenumber spectral data used in the CS reconstruction is obtained by sampling from the pre-generated k-space data set using a variable density

sampling mask [32] and stored in the host memory of the computer. This is consistent with the desired practical application in which the under-sampled k-space data will be grabbed from the camera, transferred, and stored in the RAM of the computer. The under-sampling rate is chosen to balance the image quality and input data size.

The following processing methods are compared using the same input data ($\mathbf{y}_u$): (1) GPU-MNUDFT-CS; (2) GPU-NUFFT-CS with different kernel functions (KB and Gaussian), over-sampling ratios ($R = 1.5$ and 2) and kernel widths ($W = 3$ and 5). For comparison, the data obtained by applying NUDFT to the 100% spectra (denoted as "NUDFT" in Fig. 5.5) and GPU-FFT-CS in Sec. 5.1 on the same amount of under-sampled linear wavenumber spectral data are also displayed. The same parameters τ and η are used in the CS experiments. The starting values are set as $\mathbf{x}^0 = 0$ and $\alpha^0 = 1$. These two starting values do not show obvious influence on the image quality and computation time.

5.2.3.1 Sensitivity roll-off

The point spread function (PSF) and sensitivity roll-off with different processing methods are presented in Figs. 5.5(a)–5.5(f). Here 128 A-scans each with 2048-pixel were averaged for each position. CS reconstruction uses 30% spectral data. $\tau = 3$. $\eta = 15$. As can be seen from Figs. 5.5(e) and 5.5(f), there are only subtle observation differences between the data with the same kernel but different R and W; thus, only three NUFFT-CS results are displayed: "NUFFT-CS Gaussian R2 W5", "NUFFT-CS

KB R1.5 W3", and "NUFFT-CS KB R2 W5".

The NUFFT-CS results in Figs. 5.5(d), 5.5(e), and 5.5(f) show flatter sensitivity roll-off with a smaller background noise and side-lobes at larger image depths than the FFT-CS result in Fig. 5.5(b). Compared to the NUDFT result on 100% spectral data in Fig. 5.5(a), NUFFT-CS results exhibit lower background noise, which is approximately -70dB against -60dB for NUDFT. NUFFT-CS provides a similar performance to MNUDFT-CS, as shown in Fig. 5.5(c). The NUFFT result using KB kernel with $R = 2$ and $W = 3$ is added to Figs. 5.5(g)–5.5(h) to illustrate the influence of W on the image quality. The comparison between PSFs at an imaging depth of 0.85 mm is presented in Fig. 5.5(g). The intensity of each A-scan is normalized. FFT-CS shows large side-lobes while NUFFT-CS, MNUDFT-CS and NUDFT have relatively flat background and no noticeable side-lobes. Compared to NUDFT, NUFFT-CS and MNUDFT-CS have much lower background noise. The maximum amplitude of PSFs from different methods is exhibited in Fig. 5.5(h). MNUDFT-CS and NUFFT-CS show a very close sensitivity roll-off to NUDFT while FFT-CS has much faster sensitivity attenuation as the image depth increases. Figure 5.5(i) shows the SNR versus axial position for various processing methods with the SNR definition in [82]. All CS methods have a very close SNR at the imaging depth less than 0.8 mm, while the SNR of FFT-CS becomes lower as the image depth increases. Compared to NUDFT, CS methods usually show better SNR since CS is effective in reducing noise [30, 87].

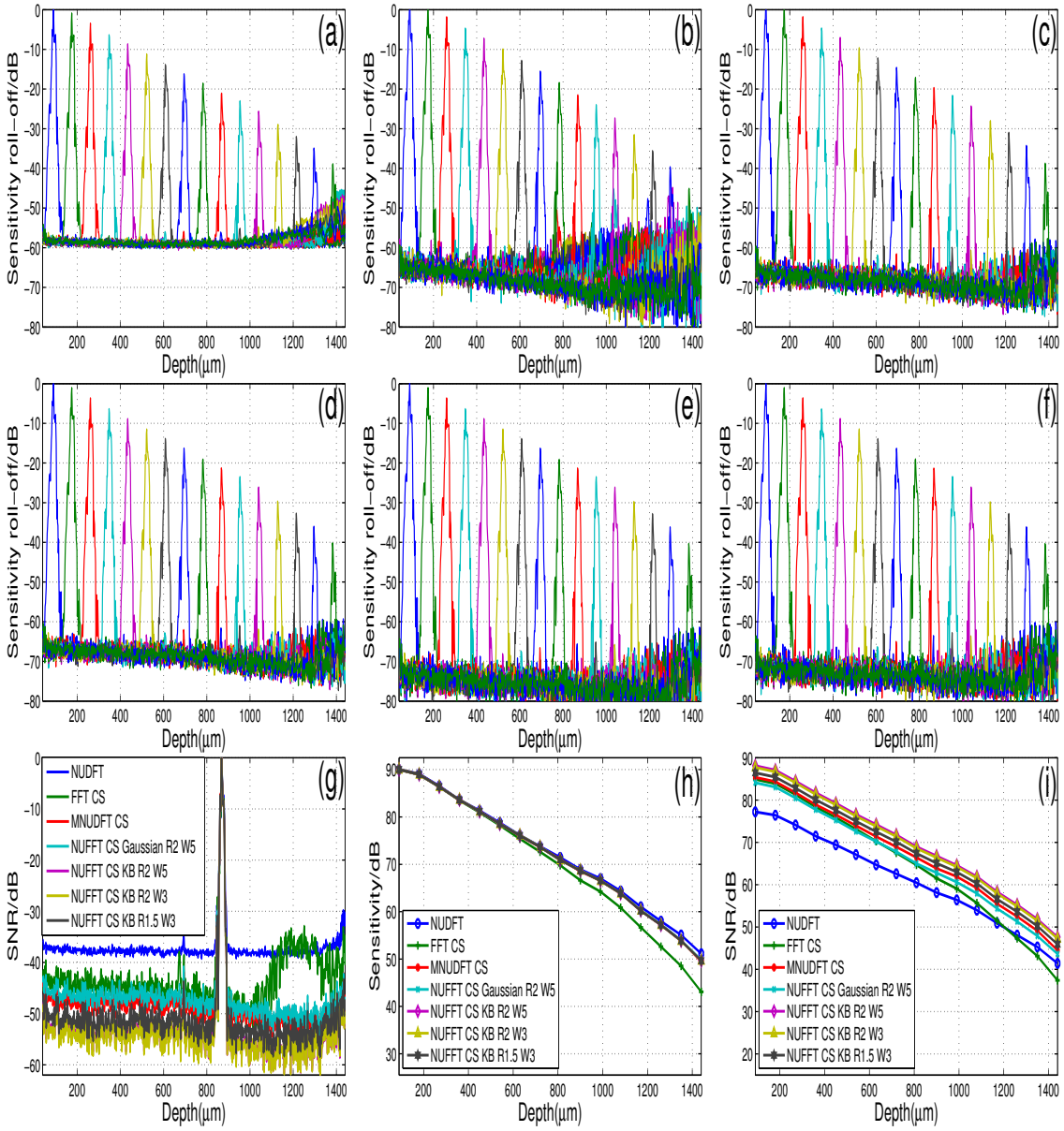

Figure 5.5: Sensitivity roll-off of different processing methods: (a) NUDFT on 100% data; (b) FFT-CS; (c) MNUDFT-CS; (d) NUFFT-CS with Gaussian kernel (R=2, W=5); (e) NUFFT-CS with KB kernel (R=2, W=5); (f) NUFFT-CS with KB kernel (R=1.5, W=3). (g) comparison of PSFs at a certain image depth using different processing methods; (h) maximum amplitude of PSFs using different processing methods; (i) SNR versus image depth for different processing methods.

All NUFFT-CS implementations reconstruct the peak of PSFs successfully. NUFFT-CS with KB kernel shows better denoising effect compared to the one with Gaussian kernel. Bigger over-sampling ratio R and kernel width W result in better suppression of the background noise level. Figures 5.5(g) and 5.5(i) show that the "NUFFT CS KB R2 W5" result is very close to the result of "NUFFT CS KB R2 W3". Both of them have a lower background noise than "NUFFT CS KB R1.5 W3" result. R shows bigger influence on the image quality than W.

5.2.3.2 Speed comparison

The averaged computation time for a single A-scan using GPU-NUFFT-CS is listed in Table 5.1. The time is averaged based on 100 runs of B-scan reconstruction. The size of B-scan is 2048(axial)×1000(lateral) pixels. As stated above, the computation time of our implementation is determined by the stopping iteration number which is set as $\eta = 15$. The pre-computed mode is tested for both the Gaussian and KB kernel while only the Gaussian kernel is evaluated in the on-the-fly mode since the modified Bessel function has not been incorporated into CUDA. In both modes, the kernel width W has more influence on the processing speed compared to the over-sampling ratio R. For comparison, the averaged computation time of GPU-MNUDFT-CS and GPU-FFT-CS is 626.67 μs and 47.2 μs, respectively.

The time for transferring the under-sampled spectral data from host to device and the image result from device to host is contained in the time listed in Table 5.1. In

Table 5.1: Computation time (μs) for an A-scan

	pre-computed mode Gaussian and KB				on-the-fly mode Gaussian			
NUFFT-CS	R=1.5		R=2		R=1.5		R=2	
	W=3	W=5	W=3	W=5	W=3	W=5	W=3	W=5
	99.17	122.27	100.03	122.87	154.47	199.92	167.15	223.84

fact, the time listed in Table 5.1 is the integration of the time for data transferring, CS reconstruction, and data conversion. We measured the data transferring time by running experiments using 50% spectral data. The time required to transfer the under-sampled spectral data of a B-scan from host to device is $346.93\mu s$ where the size of the under-sampled spectral data of one B-scan is 1.953MB (1024×1000×2 Bytes) with data type being unsigned short. The time required to transfer the reconstructed image from device to host is $314.59\mu s$ where the size of the image is also 1.953MB (1024×1000×2 Bytes) with data type being unsigned short, since only half of the B-scan needs to be displayed. The transfer time introduces a negligible latency to the system and does not affect the image result.

The comparisons on the sensitivity roll-off and computation time show that NUFFT-CS has similar performance as the MNUDFT-CS while providing a speed enhancement of more than 5 times when conducted in the pre-computed mode. Compared to FFT-CS, it shows much lower background noise and less side-lobes at the large image depths which also indicates a higher local SNR at deeper axial positions. Also, it employs a more flexible and time-saving sampling pattern than FFT-CS. Compared

to NUDFT with 100% data, NUFFT-CS has negligible difference on the sensitivity roll-off but exhibits better SNR at any image depth using only a portion of the spectral data.

In the GPU-NUFFT-CS SD OCT, KB kernel is a better choice which offers better image quality. R has bigger influence on the image quality while W impacts more on the program speed. KB kernel with $R = 2$ and $W = 3$ would be a good choice for GPU-NUFFT-CS SD OCT.

5.2.3.3 In vivo SD OCT imaging

In vivo SD OCT imaging of biological samples was conducted to confirm the performance of GPU-NUFFT-CS. The experiments were executed on the same triple-GPU architecture as in Sec. 5.1.

The SD OCT images of an orange and human skin obtained with GPU-NUFFT-CS are presented in Figs. 5.6(b) and 5.6(d), respectively. The sampling rate was 40%. $\tau = 3.0$. $\eta = 10$. The KB kernel with over-sampling ratio 2 and kernel width 3 is used in NUFFT. The reconstruction results by applying NUDFT to 100% data are shown in Figs. 5.6(a) and 5.6(c), respectively as a reference. The images of the same sample are shown in the same dynamic range. Figures 5.6(a) and 5.6(b) are the average of 100 frames while Figs. 5.6(c) and 5.6(d) are single frame results. The NUFFT-CS achieves an accurate reconstruction which is very close to the reference image. Also, smaller background noise is observed in the NUFFT-CS results. The peak signal-to-noise

ratio (PSNR) is computed for Figs. 5.6(a)–5.6(d) for a quantitative comparison, with the definition PSNR= $10\log_{10}(\max^2(f(\mathbf{x}))/\text{var})$. $f(\mathbf{x})$ is the intensity of the B-scan while var is the standard deviation of the selected background regions outlined by the white dash rectangles in Fig. 5.6. The comparison of PSNR is 73.10dB to 79.92dB for Fig. 5.6(a) to 5.6(b), respectively and 69.48dB to 76.85dB for Fig. 5.6(c) to 5.6(d), respectively. GPU-NUFFT-CS shows an average of 7dB improvement on the SNR. CS tries to maximize the sparsity of the signal in the transformed domain while preserving the data fidelity in the measurement domain. Higher sparsity indicates that a larger percentage of the signal coefficients are zero or close to zero. Thus CS has been shown to be good at reducing background noises [23, 30, 34, 87]. For the same reason, CS reconstruction also results in loses of low-intensity features of the image.

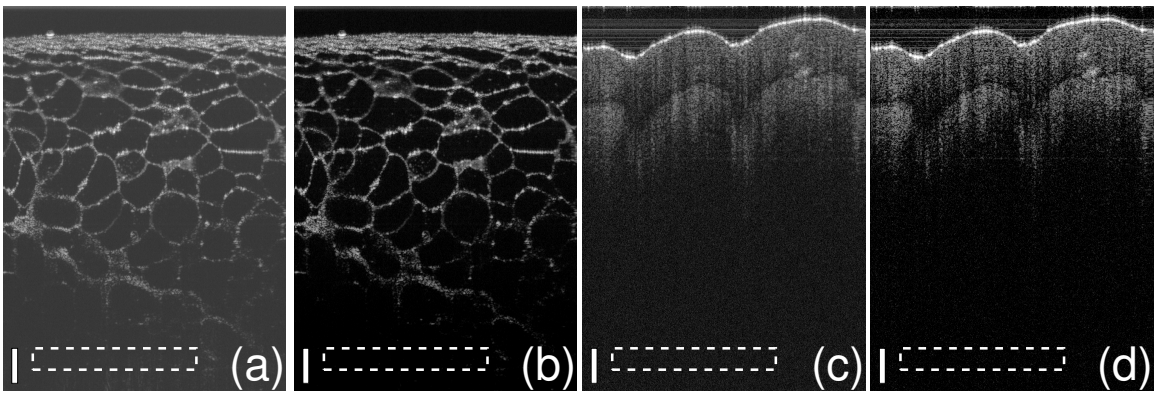

Figure 5.6: B-scans of an orange: (a) original image; (b) NUFFT-CS reconstruction result, and human skin: (c) original image; (d) NUFFT-CS reconstruction result. The scale bars represent 100 μm. The image size is 900(axial)×950(lateral).

We test the program on the realtime processing of OCT images of orange and human skin. The image frames rates in these two cases are both 72 fps with frame

size 2048×1000. The sampling rate is 40%. $\tau = 3.0$. $\eta = 10$. For comparison, GPU-MNUDFT-CS on the same architecture has a speed of 6.5 fps. KB kernel with over-sampling ratio 2 and kernel width 3 is pre-computed for NUFFT.

5.3 Realtime CS SDOCT with dispersion compensation

5.3.1 Introduction

It is well known that dispersion in SD OCT degrades the axial resolution and reduces the sensitivity of the system by introducing a frequency-dependent phase to the spectral component. The CS SDOCT with dispersion compensation has been discussed in Sec. 4.4, which demonstrates that dispersion compensation can be incorporated into CS reconstruction of SD OCT signal by adding the dispersion correcting terms to the sensing matrix. Using this modified sensing matrix, an SD OCT image with dispersion compensation can be obtained from the under-sampled spectral data. Dispersion compensation becomes a by-product of CS reconstruction; nevertheless, this method requires heavy computational loading and significantly limits the CS reconstruction speed. This is mainly due to the fact that the multiplication of the sensing matrix to the signal has to be used during the reconstruction process. Therefore, the fast Fourier transform (FFT)-based two-step acceleration proposed in Sec.

5.1 that expedites the matrix multiplication cannot be applied in this case, since the modified sensing matrix is a non-uniform discrete Fourier transformation (NUDFT) matrix.

In this section, I first demonstrated that real-time dispersion-compensated CS SD OCT imaging can be realized by multiplying the dispersion correcting terms pointwise to the under-sampled spectral data before CS reconstruction, which enables the usage of the FFT-based two-step acceleration. It is shown that as the sampling rate increases, our implementation can achieve CS reconstruction 3 to 23 times faster than that based on the modified sensing matrix.

The CS SD OCT reconstruction requires numerical operations on the matrix and vectors, which is ideal for parallel processing using the graphics processing unit (GPU). GPUs have been adapted to accelerate the CS reconstruction of various signals. In Sec. 5.1, real-time B-mode imaging at a speed of 70 fps for the GPU-accelerated CS SD OCT on a conventional desktop architecture having three GPUs.

I implemented the GPU-accelerated CS SD OCT with dispersion compensation and demonstrated real-time dispersion-compensated B-mode SD OCT imaging at excess of 70 fps with a frame size of 2048(axial) $\times$ 1000(lateral) pixels on the same triple-GPU architecture in Sec. 5.1. Experimental results show that the dispersion compensation process does not introduce an obvious burden to the CS reconstruction while it provides better axial resolution and image quality.

5.3.2 Methods

It has been described in Sec. 4.4 that dispersion compensation on full-length spectra can be realized by multiplying the dispersion correcting terms to the real spectra:

$$\mathbf{x}^d = \mathbf{F} \times (\mathbf{y}. * \Theta) \tag{5.14}$$

wher $\mathbf{x}^d$ is the dispersion compensated A-scan image. $\mathbf{F}$ is the Fourier transformation matrix. $\mathbf{y}$ is the full-length spectra. ".$*$" stands for the pointwise multiplication. Θ is the dispersion correcting vector: $\Theta = \exp(i\Phi(\omega))$; $\Phi(\omega)$ is the dispersion correcting phase vector. i is the imaginary unit. In SD OCT, compensating the second and the third order dispersion is usually sufficient, which implies $\Phi(\omega) = -a_2(\omega - \hat{\omega})^2 - a_3(\omega - \hat{\omega})^3$. ω is the vector of angular frequency corresponding to $\mathbf{y}$ and $\hat{\omega}$ is the central angular frequency. a_2 and a_3 are dispersion compensation coefficients.

Here, we demonstrated a novel method that incorporates high-speed dispersion compensation into CS SD OCT. The sensing matrix, as well as the CS reconstruction process, does not change. Thus, it is quite convenient to integrate this technique to CS SD OCT. As shown in Fig. 2, there is no observable increase in the computation time.

Denote $\mathbf{y}. * \Theta$ as $\mathbf{y}^d$. Equation 5.14 can be rewritten as:

$$\mathbf{x}^d = \mathbf{F} \times \mathbf{y}^d \tag{5.15}$$

If we have a portion of $\mathbf{y}^d$ (denoted as $\mathbf{y}_u^d$), according to CS SDOCT theory, $\mathbf{x}^d$ can be reconstructed with high accuracy by solving the following problem:

$$\mathbf{x}^d = \min_{\mathbf{x}} \|\mathbf{F}_u\mathbf{x} - \mathbf{y}_u^d\|_2^2 + \tau\|\Psi\mathbf{x}\|_1 \tag{5.16}$$

Ψ is the sparsifying operator that transforms $\mathbf{x}$ to a sparse representation, which is chosen as the identity matrix in this section, since the SD OCT image is usually sparse enough in the spatial domain. $\mathbf{y}_u^d$ can be obtained by multiplying the under-sampled dispersion correcting term Θ_u to $\mathbf{y}_u$ pointwise, where Φ_u contains the correcting terms corresponding to $\mathbf{y}_u$, which can be computed beforehand.

Thus, dispersion compensation can be incorporated into CS SD OCT by multiplying the under-sampled linear in wavenumber spectral data with a corresponding dispersion correcting term before the CS reconstruction process.

This method enables the acceleration of CS reconstruction using the FFT-based two-step algorithm, since the sensing matrix is still the uniform Fourier transform matrix. It results in significant speed enhancement.

5.3.3 Results and discussion

The proposed method is tested using spectral data from an in-house developed SD OCT system whose configuration has been described in Sec. 5.1. The data processing, as well as image rendering, was performed on the same triple-GPU architecture

described in Sec. 5.1.

The data processing flow chart for CS SD OCT with dispersion compensation on the triple-GPUs architecture is illustrated in Fig. 5.7. Each GPU was controlled by a thread that performs the CS reconstruction with dispersion compensation. Another thread displays the reconstructed B-scans in the order. To achieve enhanced acceleration and smooth the B-scan display, our programs reconstruct A-scans in one B-scan together. The NVIDIAs Compute Unified Device Architecture (CUDA) 5.5 was used to program the GPU for the general-purpose computation. The user interface was designed and programmed using C++ under the Microsoft Visual Studio 2008 environment. Compared to the data processing flow chart in Sec. 5.1, our implementation adds an extra GPU kernel for dispersion compensation that multiplies the corresponding dispersion correcting terms to the under-sampled spectral data of all A-scans in one B-scan before CS reconstruction. Thus, minor modification is needed to integrate the proposed method to CS SD OCT.

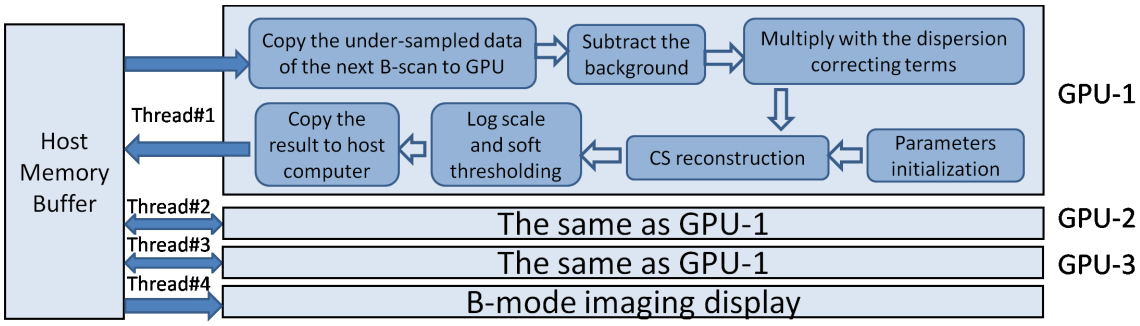

Figure 5.7: Data processing flow chart of CS SD OCT with dispersion compensation on the triple-GPU architecture.

The CS reconstruction was solved using the SpaRSA algorithm [96]. Please refer

to Sec. 5.1 for details of the implementation of SpaRSA on the GPU, as well as the influence of the CS regulation parameter τ and stopping iteration number on the image quality and computation speed.

5.3.3.1 Bench mark line rate test

We first compared the data processing speed of our method (denoted as CS-Disp-FFT) with that of CS SD OCT with dispersion compensation based on the modified sensing matrix in Sec. 4.4 (denoted as CS-Disp-Matrix), as well as that of CS SD OCT without dispersion compensation in Sec. 5.1 (denoted as CS-FFT). CS-FFT and CS-Disp-FFT are accelerated using the FFT-based two-step algorithm while CS-Disp-Matrix could not be accelerated. As stated in Sec. 5.1, the speed of our programs is independent of the sampling objects, CS regulation parameter τ, and the initial value of the CS reconstruction. The speed of CS-Disp-FFT and CS-FFT is determined by the pre-defined stopping iteration number; the speed of the CS-Disp-Matrix is determined by both the pre-defined iteration number and the sampling rate. The same stopping iteration number is used for these three methods in the experiments. The average reconstruction time for one A-scan using these three methods is displayed in Fig. 5.8(a); the benchmark line rate versus sampling rate is shown in Fig. 5.8(b). The data is averaged on 100 runs that reconstruct 250 B-scans each. The B-scan size is 2048(axial) $\times$ 1000(lateral) pixels.

As shown in Fig. 5.8, CS-Disp-FFT and CS-FFT have a relatively stable and close

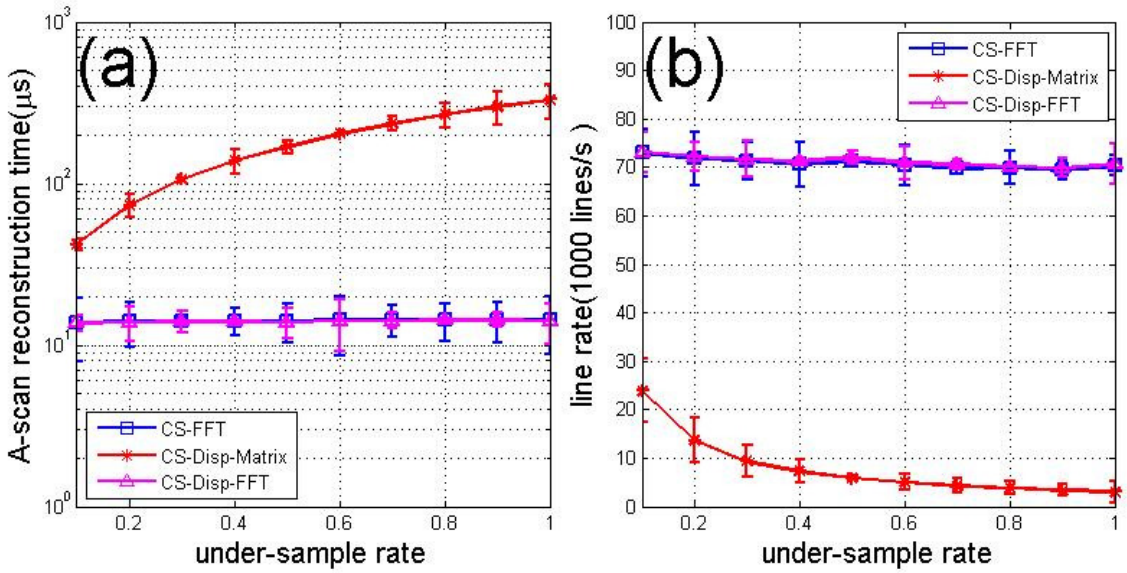

Figure 5.8: (a) Average reconstruction time of one A-scan using different methods; (b) Benchmark line rate test of different methods.

line rate, which implies that the multiplication of the dispersion correcting terms to the under-sampled k-space data adds negligible burden to the computational cost of CS SD OCT. In addition, the speed of these two programs is independent of the sampling rate. Since changing the sampling rate is unavoidable when scanning various objects, a stable speed under different sampling rate is desired in practical applications of CS SD OCT. CS-Disp-Matrix takes significantly more computation time than the other two cases under the same experimental circumstances. As the sampling rate increases, CS-Disp-FFT and CS-FFT are 3 to 23 times faster than the CS-Disp-Matrix where its speed decreases at a higher sampling rate. CS-Disp-FFT can reconstruct real-time B-mode imaging with dispersion compensation at speed in excess of 70 fps.

5.3.3.2 Point spread function comparison

The point spread functions (PSF) for the different processing methods are presented in Fig. 5.9. A mirror is used as the reference sample. The dispersion-compensated result obtained with 100% spectral measurements using the classical method [1] is added as a reference. The CS reconstructions used 30% spectral data. Here, 128 A-scans with 2048 pixels each are averaged to demonstrate the mean noise level of the PSFs. A 2cm water cell was inserted into the reference arm of the interferometer to introduce a large dispersion mismatch. Intensity of each A-scan is normalized. The same regulation parameter τ and under-sampled data are used in the CS programs. The dispersion compensation coefficients a_2 and a_3 are set as $a_2 = 400fs^2$ and $a_3 = 158fs^3$. All three CS programs obtain accurate reconstructions. The results of CS-Disp-FFT and CS-Disp-Matrix are very close to that of the classical method on 100% measurements. All of them present much sharper peaks than CS-FFT that does not incorporate dispersion compensation. The full-width-half-maximum (FWHM) of the compensated results is $4\mu m$, compared to 22 μm for CS-FFT. Fig. 3(b) shows that the result of CS-Disp-FFT is almost identical to that of CS-Disp-Matrix, while providing a significant speed enhancement. Those three CS reconstruction results show much lower background noise, which is approximately -70dB compared to -60 dB for the results obtained using the classical method. CS-Disp-FFT and CS-Disp-Matrix have lower background noise than CS-FFT. It validates that the noise reduction is a by-product of CS reconstruction.

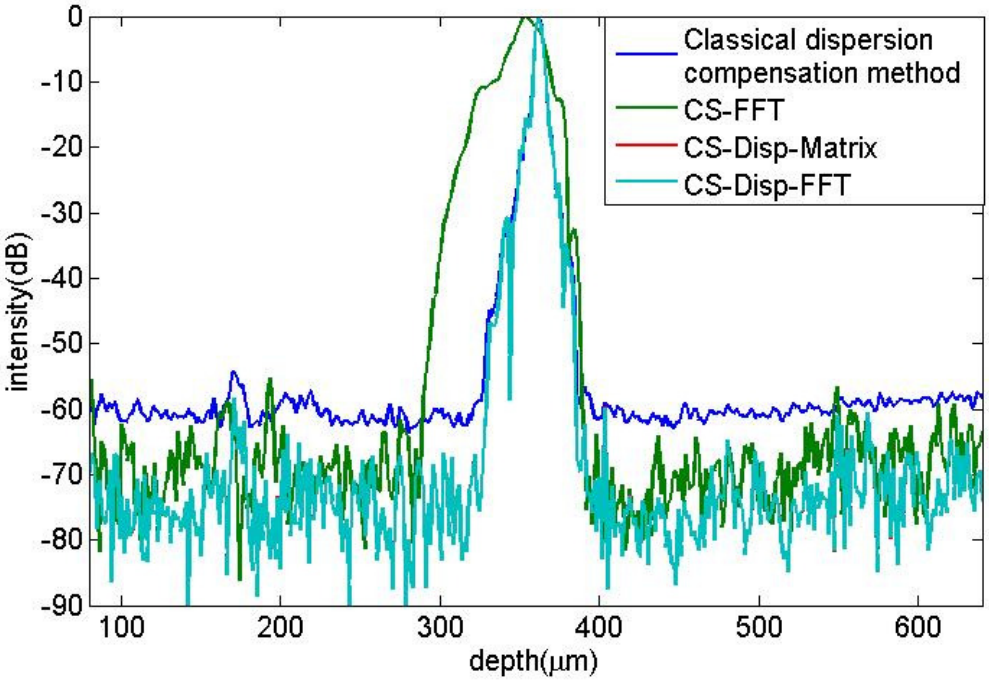

Figure 5.9: (a) Comparison of point spread functions obtained with different methods; (b) difference between the results of CS-Disp-FFT and CS-Disp-Matrix.

5.3.3.3 In vivo SD OCT imaging

In vivo SD OCT imaging on the biological samples was performed to confirm the real-time performance of the proposed method. The SD OCT images of an orange and skin of a healthy human volunteer are presented in Figs. 5.10(c) and 5.10(f), respectively. The CS-FFT results using both samples are presented in Figs. 5.10(b) and 5.10(e), respectively. The results of both samples obtained through the classical dispersion compensation method on 100% spectral measurements are also shown as a reference in Figs. 5.10(a) and 5.10(d), respectively. The sampling rate is 40% for CS reconstructions, with regulation parameter $\tau = 4.0$. Here, a 2cm water cell was

also placed in the reference arm to intentionally add big dispersion mismatch. The compensation parameters were the same as those in Sec. 5.3.3.2. The images of the same samples are displayed over the same dynamic range. As shown by the figures, CS-Disp-FFT achieves obvious dispersion correcting and its result is very close to that obtained through the classical method on 100% data. Both of them show much thinner boundary and clearer structure than the CS-FFT results. Compared to the compensated images using 100% data, CS-Disp-FFT suppresses the background noise with a darker background and better image contrast. The peak signal-to-noise ratio (PSNR) is computed for all the subfigures in Fig. 5.10 for a quantitative comparison, which is presented in Table 5.2. The definition of PSNR is:

$$PSNR = 10\log_{10}(\max^2(f(x)/var)) \tag{5.17}$$

where $f(x)$ is the intensity of the B-scan; var is the standard deviation of the intensity of the selected background regions (outlined by the white dash rectangles in Fig. 5.10). The CS-Disp-FFT has an average of 6.59dB and 2.22dB better SNR compared to CS-FFT and the classical method on 100% measurements, respectively.

Table 5.2: PSNR (dB) of the B-scans in Fig. 5.10

	Classical dispersion compensation method	CS-FFT	CS-Disp-FFT
Orange	72.28	68.83	**74.94**
Skin	97.85	92.59	**99.62**

We test the program on the realtime processing of OCT images of orange and

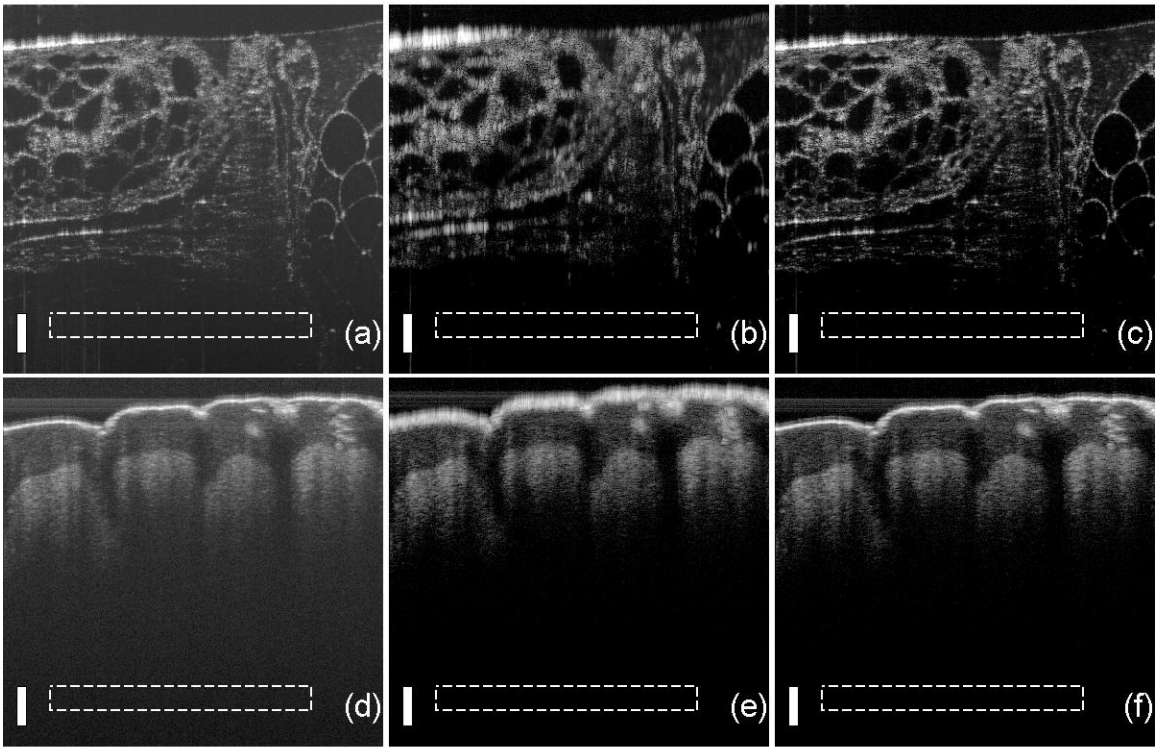

Figure 5.10: B-scan results of an orange (a-c) and human skin (d-f) with 2cm water induced dispersion using different methods: (a,d) classical dispersion compensation method on 100% data; (b,e) CS-FFT without dispersion compensation; (c,f) CS-Disp-FFT with dispersion compensation. The scale bars represent 100 μm.

human skin. The reconstruction speed of these three videos are 72 fps with frame size 2048(axial) $\times$ 1000(lateral) pixels. For a comparison of speed, the CS-Disp-Matrix program using the same spectral data operated on the same computer architecture has a speed of 6.5 fps. CS-Disp-FFT shows a clear advantage on the computation speed, while providing a close dispersion compensation effect. Compared to CS-FFT, CS-Disp-FFT exhibits a much thinner boundary and better image quality with a comparable computation speed. In addition, the above experiments show that the speed of our method is independent of the scanning object, which is especially

important for the development of practical CS-OCT systems.

The dispersion compensation coefficients in our experiments are set empirically. Their values only influence the image quality and have no impact on the reconstruction speed. We can also obtain them using a similar iterative procedure as [1], which tries to find their optimal value through an iterative procedure.

Chapter 6

Denoising and Volumetric Imaging of CS SDOCT

6.1 MoCS-OCT: Modified CS SDOCT with noise reduction

6.1.1 Introduction

Due to the interferometric nature of OCT, speckle is an inherent feature of OCT imaging. Thus specklenoise reduction has long been a focus of OCT research. Various methods for speckle noise reduction have been proposed based on hardware and software approaches to enhance the quality of OCT imaging [34,37,42–44]. One very

successful and classical method is through the averaging of a series of B-scans that are successively recorded with a slightly lateral offset [44]. Typically, averaging 10 B-scans with a 300um total lateral offset is sufficient to provide a satisfying image quality. However, this classical averaging method may require a high data acquisition speed and a large data storage space. Since CS requires only a fraction of the total data to accurately recover an image, this can significantly reduce the total data amount and increase the data acquisition speed.

Furthermore, since CS enhances signal's sparsity, it has been shown to be good at reducing noise [23]. One outstanding benefit for using CS reconstruction compared to commonly used denoising methods, such as filters, is that no additional denoising procedure is needed after the reconstruction. The noise reduction is simply a by-product of the reconstruction procedure.

Recently, Modified-CS [45,46] method was proposed to reconstruct a sparse signal with partially known support. When the known support is accurate enough and its size is small enough, sufficient conditions for exact or accurate reconstruction using Modified-CS are significantly weaker than those using CS [45]. In other words, with the knowledge of an accurate and small size support, Modified-CS has smaller error rate than CS with same sample rate without extra computation requirement. Denote the known support as $\mathbf{T}$. Modified-CS tries to find the sparest signal whose elements'

indices are out of the set $\mathbf{T}$ among all signals that satisfy the data constraint:

$$\underset{\mathbf{x}}{\text{minimize}} \ \|(\mathbf{Wx})_{\mathbf{T}^c}\|_1, \text{s.t.} \|\mathbf{F_u x} - \mathbf{y}\|_{\mathbf{2}} \leqslant \varepsilon \tag{6.1}$$

Let $\beta = \mathbf{Wx}$. $(\beta)_{\mathbf{T}}$ refers to a sub-vector of β that contains the elements with indexes in $\mathbf{T}$. $\mathbf{T}^c$ is the complement of $\mathbf{T}$, w.r.t $[1, 2 \dots n]$, n is the length of β. $b\%$ energy support $\mathbf{T}$ is defined as smallest subset of $[1, 2 \dots n]$ that satisfies $\sum_{i \in \mathbf{T}} \beta_i^2 / \sum_{j=1}^{n} \beta_j^2 \geqslant b\%$. Thus $(\beta)_{\mathbf{T}}$ prefers elements with a higher intensity. Equation (6.1) shows that Modified-CS minimizes l_1 norm of $(\mathbf{Wx})_{\mathbf{T}^c}$, not the entire signal. Hence intensity of $(\mathbf{Wx})_{\mathbf{T}}$ is not reduced. The noise is random and usually has relatively small intensity which prevents it from being in $(\mathbf{Wx})_{\mathbf{T}}$. Thus most elements in $(\mathbf{Wx})_{\mathbf{T}}$ are desired signals. With $(\mathbf{Wx})_{\mathbf{T}}$ not decreased, the ratio between the signal intensity to noise intensity becomes higher for Modified-CS. Thus it can be said that Modified-CS has a better denoising property than CS.

6.1.2 Methods

To study Modified-CS in OCT, we used data obtained from a spectral domain OCT (SD OCT) system. The details of SD OCT can be found in [30]. Relative errors when using Modified-CS and CS on an OCT image of a mirror surface is shown in Fig. 6.1. The relative error for each A-scan $\mathbf{x}$ is defined as $\|\mathbf{x} - \mathbf{x}^*\|_2 / \|\mathbf{x}^*\|_2$ where $\mathbf{x}^*$ is A-scan obtained from the fully-sampled data. The support knowledge

can be deduced from the previous A-scan if supports for adjacent A-scans are similar enough. For Modified-CS, since it lacks prior support knowledge, the first A-scan is reconstructed from 50% measurements using CS. For the second and following A-scans, sample rate is 7% and reconstruction is done using Modified-CS. In the first A-scan reconstruction, 99% energy support criteria was imposed for the reconstruction to ensure high quality reconstruction and the result is used to reconstruct second A-scan where 99% energy support was also imposed for the computation. This step was repeated for the subsequent A-scan generation until entire B-scan is reconstructed. CS uses the same sampled data and reconstruction domain as Modified-CS except all reconstructions are performed using CS. Since A-scans in SD OCT are independent of each other during both acquisition and reconstruction processes, Modified-CS and CS reconstruction is performed on each A-scan for the simplicity sake. As we can

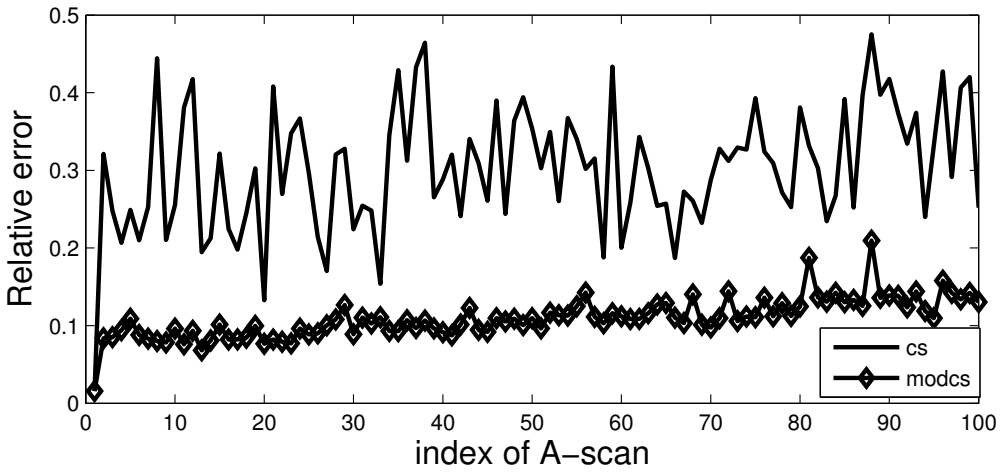

Figure 6.1: Relative error for A-scans of a mirror surface.

see from Fig. 6.1, Modified-CS has a much smaller relative error than CS. In fact,

for such a small sample rate, CS cannot provide and accurate reconstruction. It can also be inferred that supports for adjacent A-scans often change slowly. However, due to noise and complexity of OCT image, this is not always true. Thus the support knowledge obtained from previous A-scan may not be accurate enough.

The noise can be further reduced by averaging the Modified-CS reconstruction results. Modified-CS reduces the amount of data needed for averaging to obtain a desired image quality since it increases local contrast, SNR and CNR. The successive B-scans used in averaging method are recorded at nearly identical position and has the property that the supports of same position A-scans of all B-scans are similar enough, e.g. supports of the first A-scans of all B-scans are very similar. Due to duplicated denoising effect (Modified-CS + averaging), Modified-CS based averaging method reduces the noise further without influencing the signal.

Assuming the sequence of B-scans has length m and each B-scan contains l A-scans, pseudo code for Modified-CS based averaging method is as follows: The size of $\mathbf{T}_i$ used for Equation (6.1) is controlled in order to satisfy support size requirement of Modified-CS: after step 1, if size of $\mathbf{T}_i$ is too big (e.g. more than $0.1n$), it means b is set too high when computing support for this A-scan. Then $\mathbf{T}_i$ is regenerated and only contains index of elements whose intensity is among 10% largest in $\mathbf{W}\mathbf{x}_{i1}$. In this case, $(\mathbf{W}\mathbf{x}_{i1})_{\mathbf{T}_i}$ has less than b% energy. According to our experience, $(\mathbf{W}\mathbf{x}_i)_{\mathbf{T}_i}$ needs to contain at least 70% energy for a satisfying high-quality reconstruction, which is usually true for SD OCT image. This support size restriction depends on

Modified-CS-based averaging method

```
# Step 1: compute support from first B scan
for i = 1 to l do
   take full sampled measurement y_i1
   reconstruct x_i1 using y_i1 by CS or IFFT(y_i1)
   compute b% energy support, record as T_i
end for
# Step 2: reconstruct following B-scans by Modified-CS
for j = 2 to m do
   for i = 1 to l do
      take under-sampled measurement y_ij
      solve Equation (6.1) for x_ij using y_ij and T_i
   end for
end for
# Step 3: averaging result for each A-scan
for i = 1 to l do
   x̂_i = mean(x_i1 + ... + x_im)
end for
```

the available number of measurements k (it can be larger if k is larger). Restricting size of $\mathbf{T}_i$ between $0.2k$ and $0.33k$ is a good heuristic.

6.1.3 Results and discussion

Modified-CS averaging method was compared with classical averaging method and CS based averaging method. Classical averaging method was performed by first taking fully sampled measurement for all A-scans. Then inverse fast Fourier Transform (IFFT) was applied to each A-scan. Finally, B-scans were averaged using step 3 mentioned above. CS based averaging method was implemented using the similar step as the Modified-CS averaging method using the same measurements. It replaces Equation (6.1) in step 2 by Equation (3.13) and no support is computed in step 1.

Equation (3.13) and (6.1) are solved using YALL1 from [106] with default parameters.

A comparison study was performed using a chicken fat sample. 50% measurements are taken for step 2. $m = 10$; $l = 450$; $b = 85$; $\varepsilon = 60$; $\mathbf{W} = \mathbf{I}$; $\mathbf{y}$ is uniformly random sampled; The size of each reconstructed A-scan is 2048. Reconstruction results are shown in Fig. 6.2. The B-scan images are cut for better display and all four images are shown in the same dynamic range.

As we can see from Fig. 6.2, averaging results in (b),(c),(d) show that speckle noise has been reduced significantly compare to a single frame image in (a). Result from Modified-CS based averaging method shows improvement in the overall clarity of the image and the reduced noise level as shown by the fact that the background part of (d) is much darker than that of (b) and (c). The local contrast for (b) (c) (d) in Fig. 6.2 are 4.89dB, 9.52dB and 10.70dB respectively. SNR are 16.49dB, 22.47dB and 25.57dB respectively. CNR are 25.33dB, 25.57dB, 26.61dB respectively.

The definitions for local contrast, SNR and CNR are:

$$\text{local contrast} = \frac{\mu_o}{\mu_b} \tag{6.2}$$

$$\text{SNR} = 20 \times \log_{10} \left(\frac{\sqrt{\frac{1}{N_o} \sum_{(i,j)\in\text{object}} I(i,j)^2}}{\sqrt{\frac{1}{N_b} \sum_{(i,j)\in\text{background}} I(i,j)^2}} \right) \tag{6.3}$$

$$\text{CNR} = 20 \times \log_{10} \left(\frac{\mu_o - \mu_b}{\sigma_b} \right) \tag{6.4}$$

where N_o and N_b are the number of points in selected object region and background

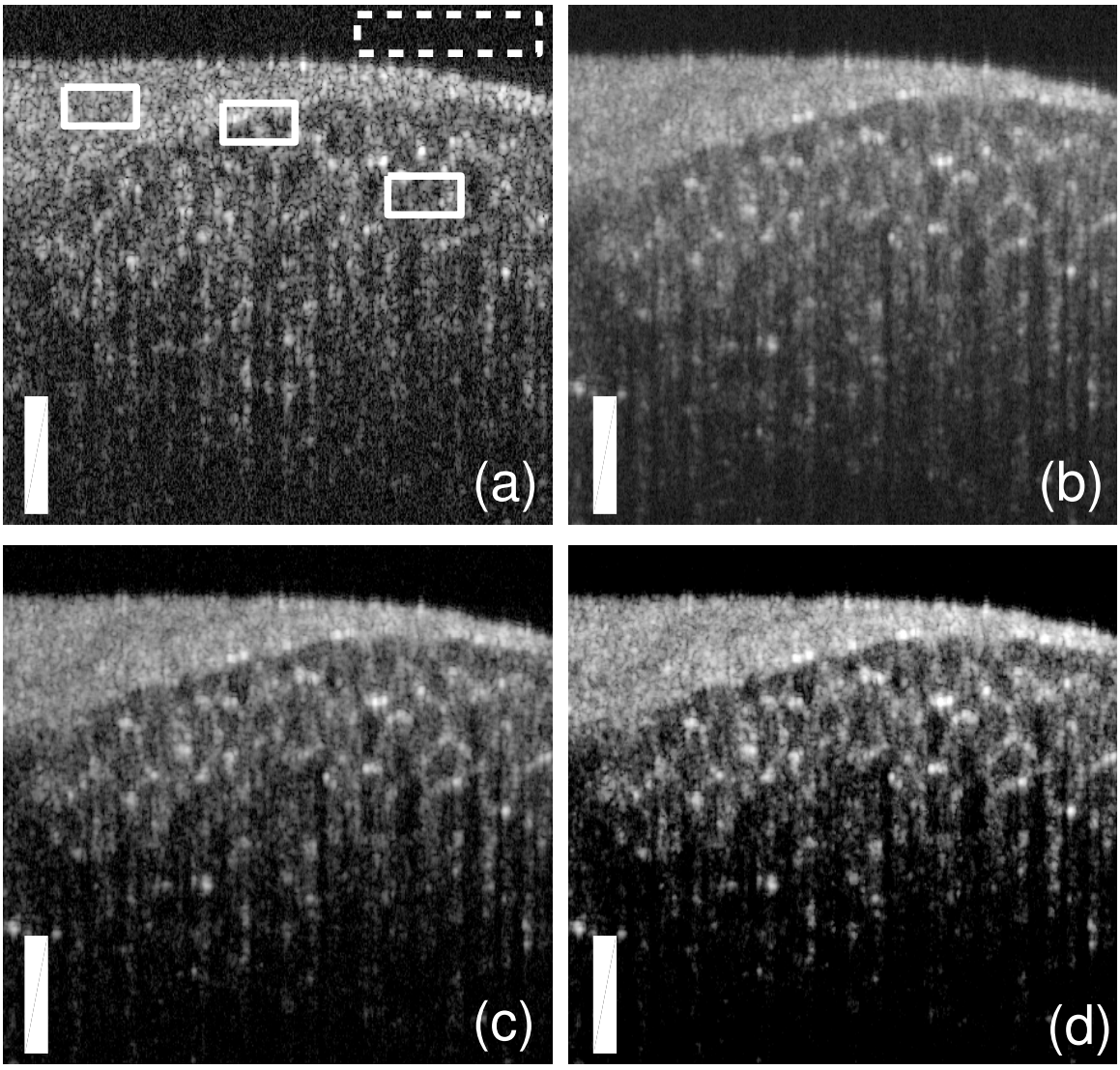

Figure 6.2: Results of chicken fat for (a) single frame (image size 320×450)(b) classical averaging result (c) CS based averaging result (d) Modified-CS based averaging result. The scale bars represent 100 μm.

region respectively. As is shown in Fig. 6.2 (a) and 6.3 (a), the areas marked within the solid rectangles are object regions of interest while the dash rectangle area represents a background region. All the selected areas have the same size. $I(i,j)$ is the intensity. μ_o and μ_b are mean of intensity of object region and background region respectively. σ_b is the standard deviation of intensity in background region.

We then compared the methods using a polymer layer phantom imaging. $m = 10$; $l = 250$; $b = 99$; $\varepsilon = \text{norm}(\mathbf{y}) * 0.025$; $\mathbf{W}$ is 4-level daubechies4 wavelet transform matrix; $\mathbf{y}$ is uniformly random sampled. The under-sample rate for step 2 is 37.5% . Reconstruction results are shown in Fig. 6.3. The B-scan images are trimmed for better display. It can be clearly seen that Modified-CS based averaging method leads to a better overall image quality. The local contrast for (b) (c) (d) in Fig. 6.3 are 5.96dB, 5.64dB and 7.79dB respectively. SNR are 17.34dB, 16.96dB and 19.88dB respectively. CNR are 26.10dB, 24.02dB and 26.40dB respectively.

The two experiments above have shown that Modified-CS based averaging method reduces the storage requirement signicantly compared to the classical averaging method while resulting in a better image quality than both classical averaging method and CS based averaging method. It is noteworthy to mention that the improvement as the result of Modified-CS reconstruction is consistent, no matter which domain the reconstruction is implemented. Whereas the improvement is not guaranteed for CS method (as is shown in Fig. 6.3). Our experimental result have shown that Modified-CS method always provides a larger improvement than CS method.

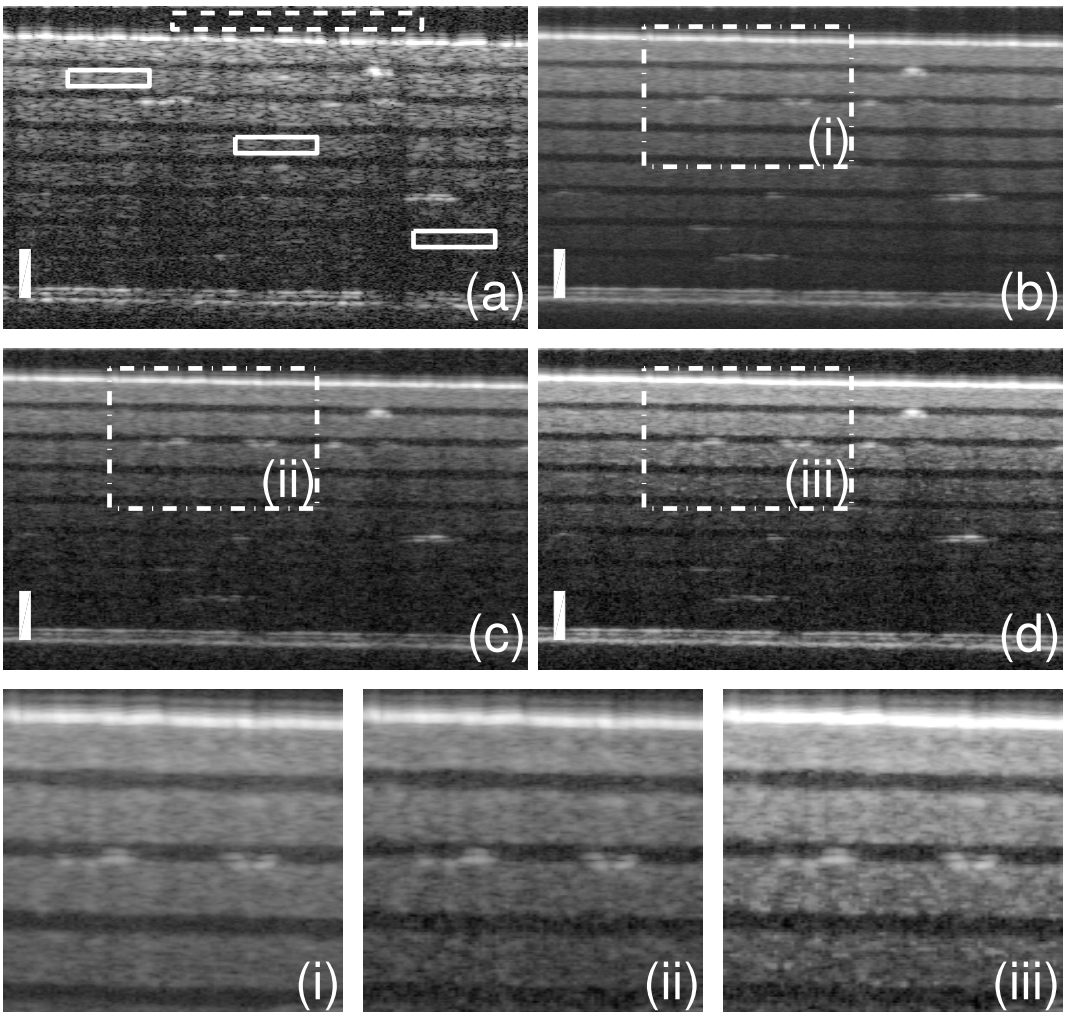

Figure 6.3: Results of polymer layer phantom imaging for (a) single frame (image size 480×250) (b) classical averaging result (c) CS based averaging result (d) Modified-CS based averaging result, (i)-(iii) are for close comparison of dash-dot rectangle regions of (b)-(d). The scale bars represent 100 μm.

For the reconstruction, ε can be estimated from the measurements when no tissue is scanned. Enlarging ε will result in better image quality while losing more small signals. Higher energy support percentage will improve image quality as long as the support size requirement is satisfied [45]. A larger sample rate gives a slightly better image while resulting in a higher storage requirement and a longer acquisition time. The reconstruction domain should be chosen in which the signal has sparsest representation. However, the image in spatial domain, where local contrast, SNR and CNR are tested, usually offers higher sparsity because the noise is reduced directly. One of the problems with the proposed method is that some small signals tend to get lost because of the sparsity enhancement, although this problem can be alleviated by averaging, it is still an open problem associated with CS relevant methods.

6.2 CS SDOCT Volumetric imaging

6.2.1 Introduction

Ultrahigh resolution 3D optical coherence tomography (OCT) imaging is a powerful and efficient technique for disease diagnosis and medical treatment of various tissues [47–50]. It is usually obtained by assembling consecutively acquired B-scans. 3D OCT provides volumetric visualization and quantification of tissue morphology features, reduction of the sampling error introduced by missing a sample location in 2D imaging, and reconstruction of en-face view of the tissue [7]. Nevertheless,

3D OCT imaging requires a long data acquisition time and large amount of spectral measurements which makes it susceptible to the unavoidable motion artifact.

Recently, compressive sensing (CS) [21, 22] has been studied extensively for the high-resolution OCT image reconstruction [30–34, 36–38, 87, 107]. It has been shown that high quality OCT image can be reconstructed using a fraction of the spectral measurements required by the Shannon/Nyqusit theory through CS reconstruction, which reduces the data acquisition time as well as the spectral data size while preserving the image quality. Application of CS on the OCT volumetric imaging has been reported in [37]. It acquires a sub set of B-scans required for the full 3D imaging and uses them to recover the missing ones through CS reconstruction. Although the under-sampled image data is used in the CS reconstruction, the sampling pattern is equivalent to sparsely sampling the spectral volume for the original 3D image in the slow-scanning lateral direction. The selected B-scans are reconstructed using 100% spectral measurements through classical method before CS reconstruction. They demonstrated volumetric OCT image recovery from up to 75% of missing data. In the studies by the same group, two similar sampling patterns are described: one consists of sparsely selected horizontal B-scans and vertical B-scans [36], while the other one takes several radial B-scans [107]. These two sampling patterns can be considered as under-sampling the original spectral volume in both the fast-scanning and slow-scanning lateral directions. The sampling patterns in [36, 37, 107] all require 100% spectral measurements for the selected B-scans. The under-sampling is

implemented on one or two lateral directions of the original spectral volume.

Most other CS OCT studies have used sampling patterns that reduce the size of spectral measurements for each A-scan, i.e. the under-sampling is done in the spectral domain directly. A-scans are reconstructed from the under-sampled spectral measurements; then several A-scans are placed together in the fast-scanning direction to form a B-scan. Volumetric image can be generated by repeating the same process for all the B-scans in the volume. Usually, 40%-50% spectral data is needed to reconstruct a high-quality OCT image for biological samples with complex morphology.

The above sampling patterns, however, only consider the under-sampling in one or two dimensions of the spectral data and use 100% measurements for the rest, which does not take full advantage of the sparse sampling, especially for the reconstruction of volumetric data. The under-sampling rate can be further reduced by exploring the sparse sampling in all three dimensions of the volumetric spectral data.

In this section, we proposed a novel sampling pattern that under-samples the spectral volume for 3D OCT image in all three directions (axial, fast-scanning lateral (B-scan) and slow-scanning lateral (C-scan)). This method can be considered as a combination of previous sampling patterns which do the under-sampling in either the axial direction or the lateral direction. The final under-sampling rate for volumetric image is the multiplication of its under-sampling rate in all three directions, which will be much lower than any one of them or the multiplication of two. Thus, our proposed sampling pattern has the potential to achieve a smaller under-sampling

rate than current methods.

To recover the 3D image, we proposed a three-step CS strategy which recovers the missing volumetric image data dimension-by-dimension. In other words, the one-dimensional CS reconstruction procedure is applied separately to the recovery of 3D image data in each direction. Our proposed sampling pattern and CS reconstruction strategy can also be easily adapted to the reconstruction of 2D OCT image to reduce the size of its spectral measurements as well as the scanning time. To the best of our knowledge, this is also the first time that under-sampling in both dimensions of the B-scan has been explored for CS OCT imaging.

Our proposed sampling pattern and CS reconstruction strategy are tested on the OCT data of in vivo biological tissues. The experimental results show that high-quality and high-resolution OCT images can be reconstructed from less than 20% of the spectral measurements.

6.2.2 Methods

The reconstruction in CS SDOCT using a subset of the measurements is obtained by solving a constrained optimization problem that maximizes the sparsity of the signal in the transformed domain while preserving the fidelity of the signal in the measurement domain. The under-sampled measurement is denoted as $\mathbf{y}_u$ and the desired image signal is denoted as $\mathbf{x}$. Both of them are one-dimensional in this paper. Note that $\mathbf{x}$ is always in the spatial domain. The CS reconstruction process can be

formulated as:

$$\min_{\mathbf{x}} |\Psi\mathbf{x}\|_1 \quad \text{s.t.} \|\mathbf{MFx} - \mathbf{y}_u\|_2^2 \leq \epsilon \tag{6.5}$$

where $\mathbf{M}$ is a sampling mask that denotes the data acquired in $\mathbf{y}_u$. $\mathbf{F}$ is the transformation between the measurements domain and the spatial domain. Ψ is the sparsifying operator that transforms $\mathbf{x}$ to a sparse representation. ϵ is the parameter that controls the fidelity of the signal in the measurement domain. $\| \bullet \|_1$ denotes the l_1-norm and $\| \bullet \|_2$ is the l_2-norm.

There are two different kinds of measurement domains used in this paper: spectral and spatial. For CS reconstruction in the axial direction (step 1 in Fig. 6.5), $\mathbf{y}_u$ is the under-sampled data in spectral domain and $\mathbf{F}$ is the Fourier transformation matrix. Ψ is chosen as the identity matrix since OCT image is shown to be sparse enough in the spatial domain [30, 31, 87]. For CS reconstruction in either of the lateral directions (step 2 and 3 in Fig. 6.5), $\mathbf{y}_u$ is the under-sampled data in spatial domain. $\mathbf{F}$ is the identity matrix. Ψ cannot be the identity matrix any more since the measurement domain cannot be the same as the sparsifying domain. Instead, two different kinds of sparsifying operator are used: the wavelet transformation and the Fourier transformation.

6.2.2.1 Under-sampling pattern

The proposed sampling pattern which under-samples the original spectral volume for the volumetric OCT image in all three directions is schematically presented in Fig.

6.4. The under-sampling in the C-scan is made up of randomly selected B-scans in the volume; a subset of the A-scans in each B-scan is randomly chosen to accomplish the under-sampling in the fast-scanning lateral direction; for each A-scan, a fraction of the k-space measurements is acquired. The comparison of the data amount between the acquired under-sampled spectral volume and the original spectral volume for 3D OCT image is also illustrated in Fig. 6.4. It shows that the spectral data obtained using the proposed sampling pattern is scaled-down in all three dimensions of the spectral measurements required by the Shannon/Nyquist rate.

The uniformly random under-sampling with constraint on a maximum gap size is applied to both lateral directions and the linear-k sampling mask proposed in [31] is used for under-sampling the k-space data in the axial resolution.

6.2.2.2 Volumetric data reconstruction

To reconstruct the 3D OCT image using the under-sampled spectral volume described in Sec. 6.2.2.1, we propose a three-step CS strategy, which is schematically illustrated in Fig. 6.5. In the first step, the selected A-scans in each B-scan are reconstructed from their under-sampled k-space measurements individually through CS reconstruction. The CS reconstructions of different A-scans are independent of each other in this process. In the second step, the CS reconstruction is applied to each selected B-scan row-by-row. The image data from the fully reconstructed A-scans in the first step are served as the under-sampled data ($\mathbf{y}_u$ in Eq. (6.5)) while those from

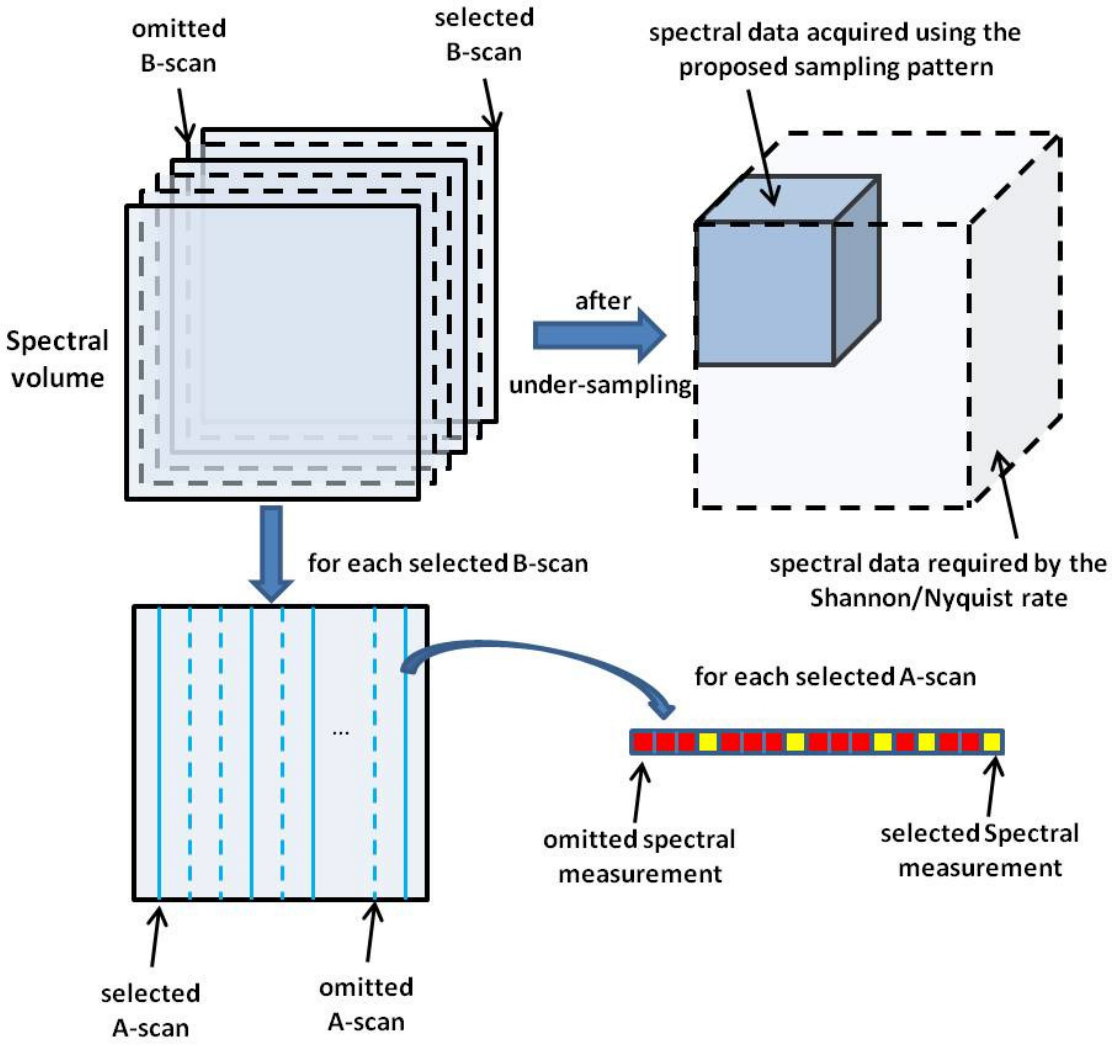

Figure 6.4: Schematic demonstration of the proposed sampling pattern which under-samples the original spectral volume in all three dimensions.

the omitted ones are recovered by CS. The reconstructions of different rows in this step are also independent of each other. After this step, all A-scans image data for the selected B-scan is recovered. In the final step, fully recovered B-scan images are used as the input data for CS volume reconstruction to recover the missing B-scan images. Note that in this stepwise process, the under-sampled measurements $\mathbf{y}_u$ in step 1 are in the k-space, while in step 2 and 3, they are image data in the spatial domain. YALL1 [106] with the default parameters is used to solve the CS reconstructions.

6.2.3 Experimental results

6.2.3.1 System configuration

The proposed sampling pattern and CS reconstruction strategy are evaluated using spectral OCT data obtained from an in-house developed SD COT. The system uses a 12-bit, 70 kHz, 2048 pixel CCD line-scan camera (EM4, e2v, USA) as the detector array of the spectrometer. The spectrometer is designed to use a superluminescent (SLED) diode light source having 105 nm bandwidth centered at 845 nm. The axial resolution of the system is approximately 4 μm in the air. Its transversal resolution is approximately 12 μm. The spectral volume used in the CS reconstructions is under-sampled from the pre-generated full size k-space volume for the 3D OCT image. The original volume data contains 250 B-scans which covers an area of 1mm$\times$1mm. The frame size of each B-scan is 2048(axial)$\times$1000(fast-scanning lateral). The camera

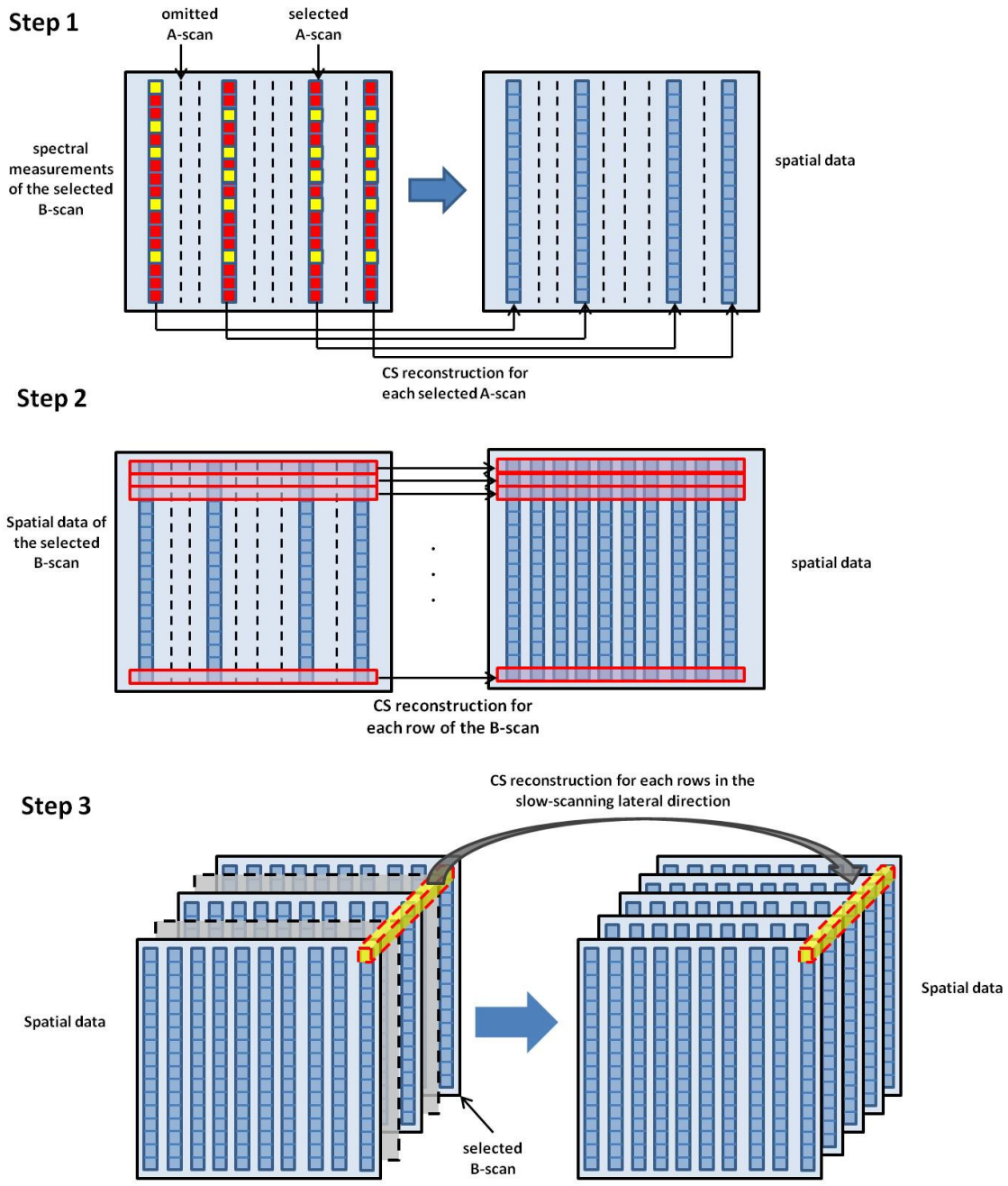

Figure 6.5: Schematic demonstration of the proposed three-step CS reconstruction strategy.

covers a spectral range of 240 nm which implies a sampling rate of 0.117 nm/pixel. All animal studies were conducted in accordance with the Johns Hopkins University Animal Care and Use Committee Guidelines.

The data processing were implemented with MATLAB R2014a on a desktop with Intel Xeon CPU (E5-2687W, 3.1GHz), 32GB RAM, windows 7 64-bit operation system. The CS programs were not optimized for speed, which can be accelerated significantly if implemented in C++ or on a graphics processing unit (GPU) [108, 109].

6.2.3.2 Image reconstruction of a B-scan of a mirror surface

We first studied the basic properties of proposed method by reconstructing a B-scan of a mirror surface. The spectral data were under-sampled in both axial and fast scanning lateral directions. As mentioned above, our proposed reconstruction strategy can be used to achieve 2D OCT image reconstruction by only taking the first two steps.

First, we fixed the overall sampling rate to 5%, which can be shown to be too small if under-sampling is implemented only in one dimension. The relative error is computed on CS reconstruction results obtained with various axial sampling rate, which changes from 5% to 100%. The corresponding sampling rate in the fast scanning lateral direction changes from 100% to 5%. The relative error is defined as:

$$e = \|f_{CS} - f_{ref}\|_2 / \|f_{ref}\|_2 \tag{6.6}$$

where f_{CS} is the CS reconstruction result; f_{ref} is the reference image. Here, we use original image as the reference, which is obtained by applying classical method to 100% spectral measurements. Two different kinds of sparsifying operators were used for CS reconstructions in the fast-scanning lateral direction: the 4-level Daubechies4 wavelet transformation and the Fourier transformation. As can be seen in Fig. 6.6(a),

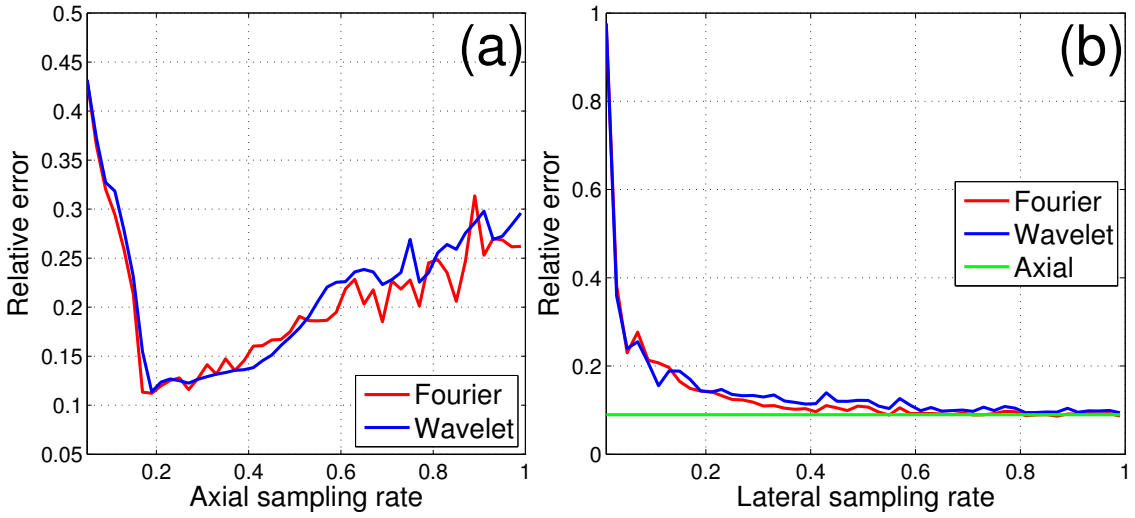

Figure 6.6: (a) relative error vs. axial sampling rate. The overall sampling rate is fixed; (b) relative error vs. lateral sampling rate. The axial sampling rate is fixed.

the relative error is high with small axial sampling rate, although the corresponding sampling rate in the fast scanning lateral direction is 100%. It is because the axial reconstruction result is used as the input data of CS reconstruction in the fast scanning lateral direction. The under-sampling rate in the fast scanning lateral direction is redudant and does not help the overall recontruction. As the axial sampling rate increases, the relative error decreases until it reaches the minimum. It shows that better reconstruction result can be achieved by exploring under-sampling in both

dimensions of the B-scan. Then the relative error increases again which implies that the sampling rate in the fast scanning lateral direction becomes too small and the axial sampling rate becomes more than enough. The relative error reaches its minimum with 20% axial sampling rate and 25% lateral sampling rate. Fourier and wavelet transformations show similar performance.

Fig. 6.6(b) shows the relative error computed on CS reconstruction results obtained using various lateral sampling rate and fixed axial sampling rate (20%). Here, the relative error is computed only on the missing A-scans, i.e. there is no spectral data obtained for these A-scans. The relative error decreases for higher lateral sampling rate and converges to the relative error level of axial CS reconstruction, which is computed on the axial reconstruction results and does not change because of fixed axial sampling rate. Similar conclusion can be drawn from comparison on CS reconstruction results with fixed lateral sampling rate and changing axial sampling rate.

6.2.3.3 Image reconstruction of a B-scan of human skin

We then evaluated the proposed methods through the reconstruction of a B-scan of human skin. Its spectral data set was under-sampled in both axial and fast-scanning lateral direction. The under-sampling rates in both directions were 50%. Both Fourier transformation and wavelet transformation was tested as the sparsifying operator in the lateral CS reconstruction. The reconstruction results are shown in Figs. 6.7(c)

and 6.7(d) respectively. The original image obtained using 100% spectral sampling in both directions is displayed in Fig. 6.7(a) as a reference. The CS reconstruction result obtained using the spectral data under-sampled only in the axial direction is shown in Fig. 6.7(b) as a comparison. The CS reconstruction for each A-scan in Fig. 6.7(b) uses 25% of the spectral measurements and the resulting A-scans are concatenated to form the B-scan. Thus, Figs. 6.7(b)-6.7(d) use the same amount of under-sampled spectral data. All the images are shown in the same dynamic range.

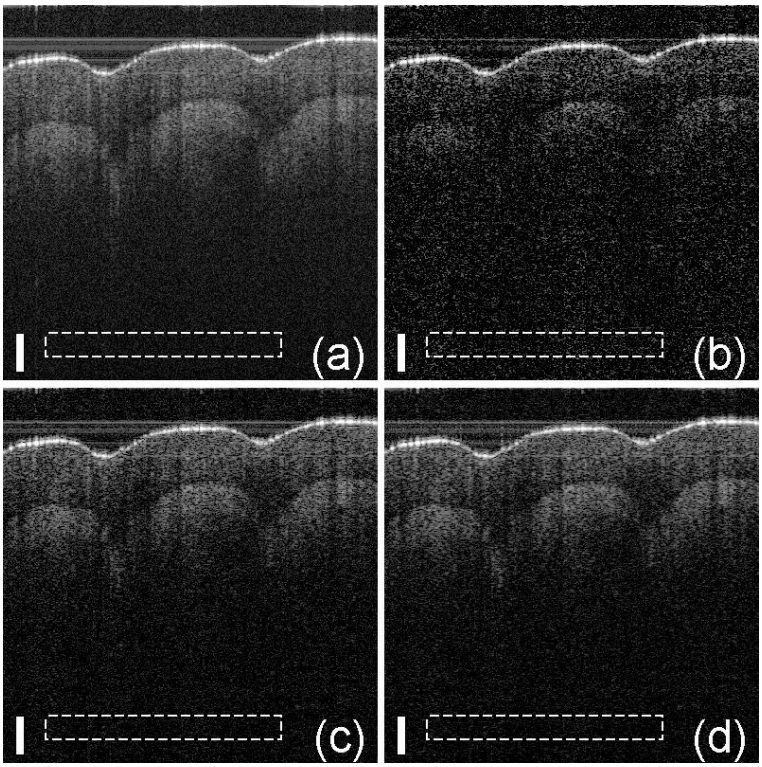

Figure 6.7: Reconstruction results of human skin. (a) is obtained using 100% spectral data. (b) is obtained using 25% spectral data for each A-scan and no under-sampling is applied in the fast-scanning direction. (c) and (d) are obtained using the proposed method, with the wavelet transformation and Fourier transformation as sparsifying operators, respectively. The under-sampling rates for the axial direction and fast-scanning direction are both 50%. The scale bars represent 100 μm. The image size in pixel is 900×925.

As can be seen from Fig. 6.7, our proposed method achieved accurate reconstructions that are very similar to the reference image. In contrast, the CS reconstruction using the same amount of spectral data that is under-sampled only in the axial direction shows obvious reconstruction error and information loss, especially for the fine structures at large imaging depths. This is mainly because 25% spectral data is not enough for obtaining an accurate CS reconstruction of an A-scan with complex morphology. Instead, our proposed method takes 50% spectral data for each selected A-scan. Thus these A-scans can be reconstructed with high accuracy. Then the missing A-scans are recovered by CS reconstruction row-by-row using the image data from these accurately recovered A-scans. The under-sampling rate for the input data used in each CS reconstruction of the fast-scanning direction is also 50%, which is usually high enough for an accurate reconstruction. Thus, our proposed method can significantly reduce the overall sampling rate by under-sampling in all dimensions, while achieving a high-quality reconstruction. The peak signal-to-noise ratio (PSNR) is computed for Figs. 6.7 (a) to 6.7(d) for a quantitative comparison of the image quality, where PSNR is defined as:

$$PSNR = 10 \log_{10}(\max^2(f(\mathbf{x}))/\text{var}) \tag{6.7}$$

$f(\mathbf{x})$ is the intensity of the B-scan. var is the standard deviation of the selected background regions that are outlined by the white dashed rectangles in Figs. 6.7(a)-

6.7(d). The PSNR for Figs. 6.7(a)-6.7(d) is 72.60dB, 69.92dB, 75.49dB, and 75.57dB, respectively. Figure 6.7(c) achieves 2.89dB and 5.57dB better PSNR than Fig. 6.7(a) and Fig. 6.7(b), respectively, while the PSNR improvements are 2.97dB and 5.65dB for Fig. 6.7(d) against Figs. 6.7(a) and 6.7(b), respectively. The CS reconstruction tries to enhance the sparsity of the signal in the sparsifying domain while preserving the fidelity of the measurements and has long been recognized to be good at reducing background noise [30, 34, 87]. For the same reason, the CS results tend to lose low-intensity features in the sparsifying domain.

6.2.3.4 Volumetric image reconstruction

We then evaluated the proposed method using a volumetric OCT image. The tested spectral volume is under-sampled from the fully acquired spectral data in all three dimensions using the proposed sampling pattern. Then the three-step strategy in Sec. 6.2.2.2 was applied to the under-sampled spectral data to reconstruct the volumetric image.

The experiment was implemented using a 3D OCT data set of a mouse cornea. The under-sampling rates of 50%, 60% and 60% are applied to the axial direction, fast-scanning lateral (B-scan) direction and slow-scanning lateral (C-scan) direction of the original 3D spectral volume, respectively. The Fourier transformation and the 4-level Daubechies4 wavelet transformation are tested separately as the sparsifying operator. The same sparsifying operator is used in the second and third steps of the

CS reconstructions in one experiment. More specifically, the images in the second columns of Figs. 6.8, 6.9 and 6.10 are obtained with wavelet transformation as the sparsifying operator in both step 2 and step 3 during the reconstruction, while the Fourier transformation is used for reconstructing the images in the third columns of these figures. The reconstructed images are post-processed for volumetric visualization using the free software ImageJ.

Figure 6.8 shows a ray-casting view of the mouse cornea. The sub images in the first row show the isotropic view of the cornea sample and those in the second row are the orthogonal slices display of the volume which show its inner structure. The first column in Fig. 6.8 presents the original 3D images as a reference that was obtained from 100% spectral measurements. The images in the second column are the CS reconstruction results using wavelet transformation, while those in the third column are the CS reconstruction results using Fourier transformation. As can be seen, CS reconstruction results using both the Fourier transformation and the wavelet transformation are very close to the reference image that uses 100% spectral measurements. The anatomical structures such as cornea, lens, and iris can be clearly visualized and they are marked out in the image.

Figure 6.9 shows representative slices extracted from the 3D mouse cornea image in different directions. Two slices are selected in each direction. The reference images using 100% spectral measurements are shown in the first column. The positions of selected slices are marked out in the reference images. The slices from 3D image

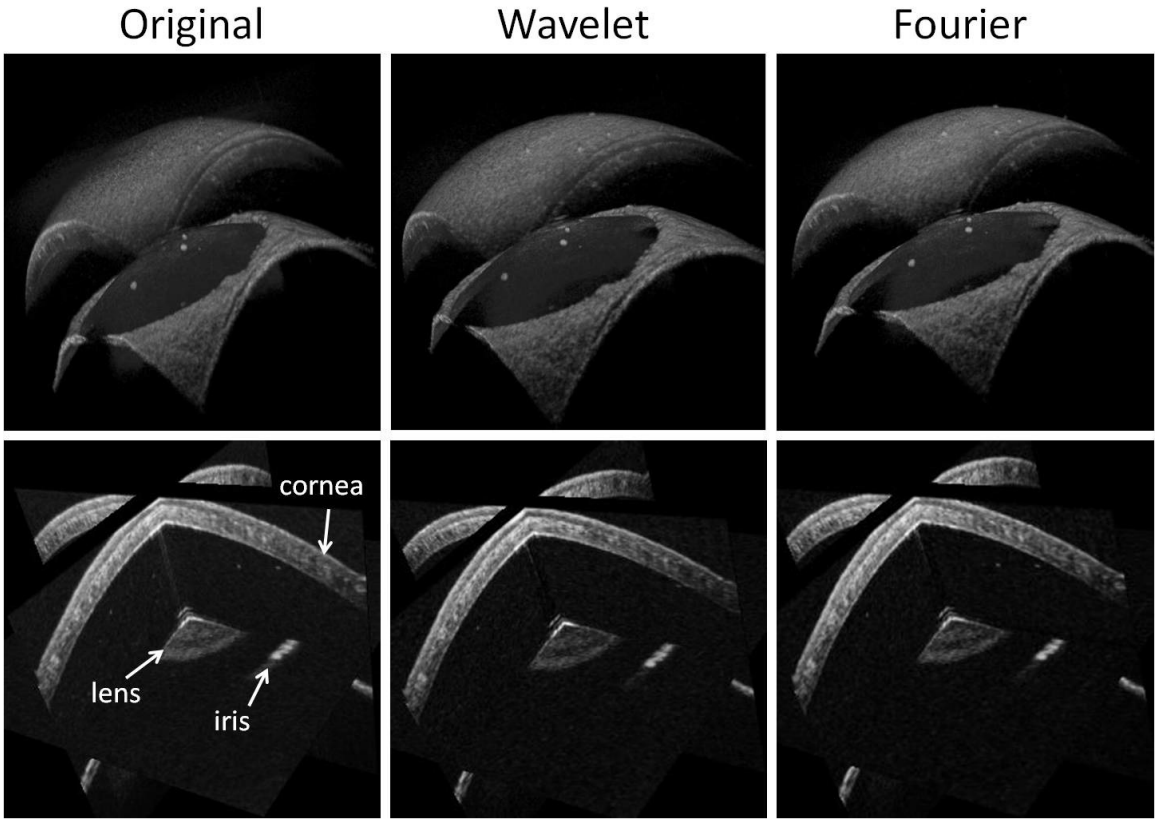

Figure 6.8: First row: volumetric visualization by ray-casting; second row: orthogonal cross-sectional display; first column: image obtained with 100% spectral data; second column: image obtained using the proposed method with the wavelet transformation; third column: image obtained using the proposed method with the Fourier transformation.

obtained using the wavelet transformation and the Fourier transformation are shown in the second and third column, respectively. To demonstrate that proposed method can recover the missing image data, the slices in both lateral directions shown in Fig. 6.9 are from the under-sampled missing data, i.e. there is no spectral data obtained for these slices and they are recovered by proposed reconstruction method. As can be seen, the missing image data are reconstructed with high accuracy which is nearly indistinguishable from the original image using 100% of the spectral volume.

One representative en-face slice of the reconstruction result is shown in Fig. 6.10 to closely examine the images obtained using different spasifying operators: the wavelet transformation and the Fourier transformation. The green rectangle areas in the first row of Fig. 6.10 are magnified in the second row. As can be seen, the difference between the results obtained using Fourier transformation and wavelet transformation is small. The wavelet transformation result achieves a better overall image quality with a sharper structure, while the Fourier transformation result exhibits a smoothed image. The smoothness is mainly because the signal in the Fourier domain has low intensity at high frequency positions, which tend to get suppressed by the sparisty enhancement during CS reconstruction. This process is similar to that of a low-pass filtering. Based on our experience, the wavelet transformation usually achieves a better image quality. However, the Fourier transformation offers a better balance in terms of achieving high image quality and fast reconstruction speed since a fast implementation through the fast Fourier transformation (FFT) is available. The

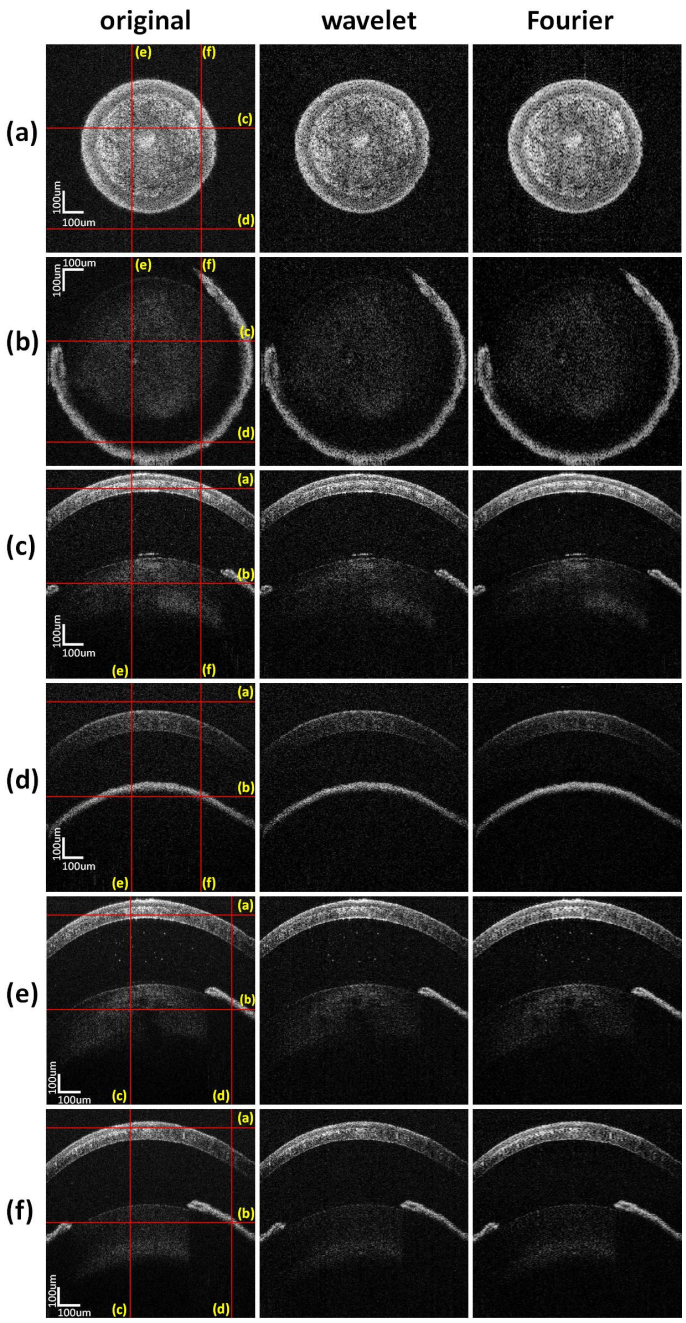

Figure 6.9: Recovered slices in the reconstructed volumetric images. Rows (a) and (b) are two en-face slices at the position of 160 μm and 800 μm below the surface, respectively. The image size in pixel is 925×250. Rows (c) and (d) are slices in the slow-scanning direction. The image size in pixel is 900×250. Rows (e) and (f) are B-scans in the fast-scanning direction. The image size in pixel is 900×925. The first column is the image obtained using 100% data. The second and third columns are images obtained using the proposed method with the wavelet and the Fourier transformation, respectively.

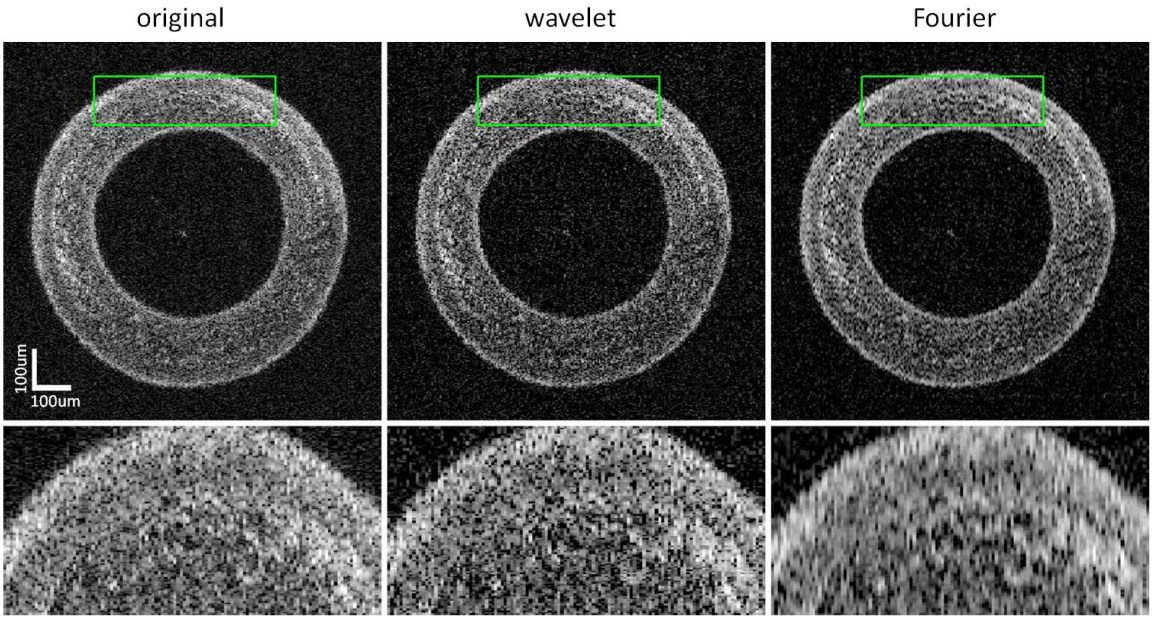

Figure 6.10: First row: representative slices obtained using 100% spectral measurements (first column) and under-sampled spectral volume using the wavelet transformation (second column) and the Fourier transformation (third column). Second row: zoom-in of the green rectangle areas in the first row.

quantitative evaluation of different sparsifying transformations is still under-study.

6.2.4 Discussion

In the proposed three-step reconstruction strategy, the order of the second and the third step can be interchanged, since the CS reconstructions in each step uses only 1D input data and there is no overlap between the input and output data of these two steps. These two steps are independent of each other. Therefore, the change of their order does not alter the reconstruction result as long as an accurate reconstruction can be achieved in each step.

In the second step of proposed reconstruction strategy, the CS reconstruction only

needs to be applied to half of the rows of each selected B-scans (rows here indicate the 1D image data in the fast-scanning lateral direction), since in regular SD OCT, only the first half of the B-scan will be displayed. This is also true for the third step of the reconstruction strategy: only the first half (i.e. top half in Fig. 6.5) of the volumetric image data needs to be reconstructed. In addition, the number of the rows that needs to be incorporated in step 2 can even be reduced further for the samples having small imaging depth. Reducing the row number will save the reconstruction time and alleviate the system memory required.

Our proposed reconstruction strategy applies 1D CS reconstruction separately to the volumetric data in each direction: in step 1, it is applied to the under-sampled spectral data of each chosen A-scan of the selected B-scan; in step 2, it is applied to each row of the selected B-scan; in step 3, it is applied to each row in the slow-scanning lateral direction of the 3D image. Although 3D transformation has also been used in the CS reconstruction of volumetric OCT image [36, 37, 107], there are several benefits for our decomposition-based strategy:

1. It is very easy to change parameters when reconstructing various data with different sparsity and noise level. Usually, a bigger ϵ is desired when reconstructing the rows corresponding to small imaging depths to reduce the noise, while a smaller ϵ is needed for those with large imaging depths to preserve the low-intensity features. It also provides the developer more choices when tuning the parameters in reconstructing the OCT data with complex morphology.

2. It saves the memory needed during the reconstruction. For instance, the full volumetric image used in the experiments has 2048×1000×250 voxels. The data size is approximately 3.81 GB assuming the data type is double (default data type in MATLAB). It is even bigger (7.63GB) when the Fourier transformation is used because the data type comes to complex double. Usually, more than one copy of the volumetric data is stored in the RAM during the CS reconstruction to represent the interim data for the transformation between the sparsity domain and the image domain or between the measurement domain and the spatial domain. Using 3D transformation will place a high burden on the memory usage. This situation will deteriorate when implementing the acceleration using GPU, since most GPUs do not have enough on-chip memory to hold more than one such big array. Complicated data decomposition is necessary if 3D transformation is used. In comparison, assuming the under-sampling rate is 50% in all dimensions and data type double, the 3D CS OCT data size in the proposed method is 954 MB in step 1 and step 2, 1.91GB in step 3. These numbers are doubled assuming the data type is complex double with the same under-sampling rate. It can be seen that our algorithm requires less memory. In addition, all the CS reconstructions in our algorithms use 1D under-sampled data. Thus, the size of interim data is very small. Our algorithm can be easily implemented on most regular GPUs.

3. It is easier to implement the GPU acceleration using proposed CS reconstruction

method. Besides the memory advantage mentioned above, it is straightforward to accelerate the computation by applying data decomposition to the volumetric data since the 1D CS reconstructions in each step are independent of each other. A large number of 1D CS reconstructions can be executed simultaneously, given that the memory of the GPU is not used up. GPU favors a big number of simple computations, instead of a small number of complex computations. We can also use more than one GPU to achieve bigger acceleration. Examples of accelerating multiple independent CS reconstructions of SD OCT data on a triple-GPUs architecture can be found in [108–110].

CS tries to maximize the sparsity of the signal in the sparsifying domain. Higher sparsity indicates that more signal tend to be zero or close to zero. Thus CS has been shown to be good at reducing the background noise. For the same reason, it usually results in the loss of low-intensity features. For the Fourier transformation case, this will smooth the image while the reconstruction result using wavelet transformation often loses low-intensity image details outright.

Many literatures have reported accurate CS reconstruction of an A-scan using 40%-50% spectral data [30,31,38,87,108,109]. In addition, the literatures [36,37,107] demonstrate volumetric CS reconstruction from a high percentage of missing data in one or both of the lateral directions. However, we found that the under-sampling rate cannot be too small for all three directions. In other words, we have not noticed satisfied reconstruction result when applying the proposed method to spectral volume

with less than 50% under-sampling rate in all three directions. The main problem is the loss of low-intensity features and details in the spasifying domain. There are mainly two reasons for this: first, CS reconstruction tends to lose low-intensity features as mentioned earlier. Recovering these low-intensity features accurately requires a very high sampling rate. Reducing the sampling rate will do more harm to low-intensity features than those high-intensity ones. Secondly, the proposed method uses three round of CS reconstruction that will also jeopardize low-intensity features in the sparsifying domain. It will require accurate enough reconstruction in every round of CS reconstruction to preserve these low-intensity features.

A bigger sampling rate results in a better image quality but increases the size of the spectral measurements as well as the data acquisition time. The optimal sampling rate for each direction is case-specific. Based on our experiences so far, setting each of them to 50%-60% is usually enough to ensure a high quality reconstruction.

Since the CS reconstruction results in axial direction (step 1) is used as input data for CS reconstructions in lateral directions (step 2 and 3), its accuracy determines the minimum error in lateral directions. As can be seen from Fig. 6.6(b), the relative error converges to the relative error level of the axial CS reconstruction as the lateral sampling rate increases. CS reconstruction in lateral direction introduces more error with smaller lateral sampling rate.

Another interesting observation is that in Fig. 6.6(a), given a fixed overall sampling rate, a small axial sampling rate results in a bigger relative error than the same

lateral sampling rate. For example, 5% axial and 100% lateral sampling rate results in a bigger error than 100% axial and 5% lateral sampling rate. Thus, it is more worthy to implement CS method in the lateral directions, given a limited computational resource.

Chapter 7

Conclusion

In this chapter, I will first summarize the works in this dissertation; then discuss some possible future work.

7.1 Summary of contributions

In this dissertation, advanced compressive sensing (CS) spectral domain optical coherence tomography (SD OCT) techniques were studied and evaluated. Several real-time implementations of CS SDOCT have also been investigated and developed.

In the first step, I proposed a novel CS SDOCT method based on the modified non-uniform discrete Fourier transform (MNUDFT) matrix which can be applied to the non-linear wavenumber spectral sampling, instead of the linear wavenumber sampling in traditional CS SDOCT. Thus the k-space grid filling and k-linear mask calibration

which were designed to obtain the under-sampled linear wavenumber sampling from the non-linear wavenumber spectra are no longer needed. In addition, dispersion compensation by multiplying the correcting terms directly to the non-linear wavenumber real spectra is proposed and incorporated into the sensing matrix of CS. Dispersion compensation is studied for the first time in the field of CS SDOCT.

To solve the bottleneck issues associated with the computation speed, I proposed to accelerate the reconstruction of CS SD OCT using the graphics processing unit (GPU). Several realtime CS SDOCT systems were implemented on a conventional desktop computer architecture having three graphics processing unit (GPU). The B-mode imaging was realized at excess of 70 fps with the under-sampled linear wavenumber sampling. The GPU-accelerated non-uniform fast Fourier transform (NUFFT)-based CS SDOCT was also implemented, which achieved more than 30 fps using the under-sampled nonlinear wavenumber sampling. In addition, I proposed a novel dispersion compensation method that enables real-time CS SDOCT image reconstruction and demonstrated that high-quality SDOCT imaging with dispersion compensation can be achieved at the speed in excess of 70 fps.

I also studied the noise reduction of SDOCT using Modified-CS-based averaging method. Compared to the classical averaging method that requires a large amount of measurements and a long data acquisition time, the averaged Modified-CS results in a significant reduction of the data amount and the acquisition time while achieving a better image quality with a much lower noise level. Finally, I proposed a novel under-

sampling pattern and reconstruction strategy for the volumetric image reconstruction using CS SDOCT. The under-sampling is done in all three dimensions of the spectral data, which further reduces the data amount. This method can be easily adapted to the CS reconstruction of a B-scan.

7.2 Future work

In the future, I look forward to expand this research in the following directions. First, I plan to build a SD OCT system that realize random sampling during the acquisition, real-time B-mode and C-mode image reconstruction, and various image rendering. This system will be tested in various clinical applications such as the surgery with OCT guidance and microvascular anastomosis guidance and evaluation. I also plan to perform an in-deep study on the optimal value of the regularization parameter in CS and the optimal sampling rate.

Bibliography

[1] M. Wojtkowski, V. Srinivasan, T. Ko, J. Fujimoto, A. Kowalczyk, and J. Duker, "Ultrahigh-resolution, high-speed, Fourier domain optical coherence tomography and methods for dispersion compensation," *Opt. Express*, vol. 12, no. 11, pp. 2404–2422, 2004.

[2] D. Huang, E. Swanson, C. Lin, J. Schuman, W. Stinson, W. Chang, M. Hee, T. Flotte, K. Gregory, C. Puliafito, and J. Fujimoto, "Optical coherence tomography," *Science*, vol. 254, pp. 1178–1181, 1991.

[3] W. Drexler and J. G. Fujimoto, *Optical Coherence Tomography: Technology and Applications*. Springer,Berlin, Germany, 2008.

[4] A. Fercher, W. Drexler, C. Hitzenberger, and T. Lasser, "Optical coherence tomography - principles and applications," *Rep. Prog. Phys.*, vol. 66, no. 2, pp. 239–303, 2003.

[5] W. Drexler, U. Morgner, R. Ghanta, F. Kartner, J. Schuman, and J. Fujimoto,

"Ultrahigh-resolution ophthalmic optical coherence tomography," *Nat. Med.*, vol. 7, no. 4, pp. 502–507, 2001.

[6] T. Ko, J. Fujimoto, and J. Duker, "Comparison of ultrahigh- and standard-resolution optical coherence tomography for imaging macular hole pathology and repair," *Ophthalmology*, vol. 111, pp. 2033–2043, 2004.

[7] W. Wieser, B. Biedermann, T. Klein, C. Eigenwillig, and R. Huber, "Multi-megahertz OCT: high quality 3D imaging at 20 million A-scans and 4.5 GVoxels per second," *Opt. Express*, vol. 18, no. 14, pp. 14 685–14 704, 2010.

[8] A. Fercher, C. Hitzenberger, G. Kamp, and S. Elzaiat, "Measurement of intraocular distances by backscattering spectral interferometry," *Opt. Commun.*, vol. 117, no. 1, pp. 43–48, 1995.

[9] G. Hausler and M. Linduer, "Coherence radar and spectral radar-new tools for dermatological diagnosis," *J. Biomed. Opt.*, vol. 3, no. 1, pp. 21–32, 1998.

[10] M. Wojtkowski, R. Leitgeb, A. Kowalczyk, T. Bajraszeski, and A. Fercher, "In vivo human retinal imaging by Fourier domain optical coherence tomography," *J. Biomed. Opt.*, vol. 7, no. 3, pp. 457–463, 2002.

[11] R. Leitgeb, C. Hitzenberger, and A. Fercher, "Performance of Fourier domain vs. time domain optical coherence tomography," *Opt. Express*, vol. 11, no. 8, pp. 889–894, 2003.

[12] J. de Boer, B. Cense, B. Park, M. Pierce, G. Tearney, and B. Bouma, "Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography," *Opt. Express*, vol. 28, no. 21, pp. 2067–2069, 2003.

[13] M. Choma, M. Sarunic, C. Yang, and J. Izatt, "Sensitivity advantage of swept source and Fourier domain optical coherence tomography," *Opt. Express*, vol. 11, no. 18, pp. 2183–2189, 2003.

[14] M. Wojtkowski, T. Bajraszeski, P. Targowski, and A. Kowalczyk, "Real-time in vivo imaging by high-speed spectral domain optical coherence tomography," *Opt. Lett.*, vol. 11, no. 25, pp. 1745–1747, 2003.

[15] S. Yun, G. Tearney, J. de Boer, N. Iftimia, and B. Bouma, "High-speed optical frequency-domain imaging," *Opt. Express*, vol. 11, no. 22, pp. 2953–2963, 2003.

[16] N. Nassif, B. Cense, B. Park, S. Yun, C. Chen, B. Bouma, G. Tearney, and J. de Boer, "In vivo human retinal imaging by ultrahigh-speed spectral domain optical coherence tomography," *Opt. Express*, vol. 29, no. 5, pp. 480–482, 2004.

[17] S. Chinn, E. Swanson, and J. Fujimoto, "Optical coherence tomography using a frequency-tunable optical source," *Opt. Lett.*, vol. 22, no. 5, pp. 340–342, 1997.

[18] B. Golubovic, B. Bouma, G. Tearney, and J. Fujimoto, "Optical frequency-domain reflectometry using rapid wavelength tuning of a cr/sup 4+/:forsterite laser," *Opt. Lett.*, vol. 22, no. 22, pp. 1704–1706, 1997.

[19] K. Zhang and J. Kang, "Real-time 4D signal processing and visualization using graphics processing unit on a regular nonlinear-k Fourier-domain OCT system," *Opt. Express*, vol. 18, no. 11, pp. 11 772–11 784, 2010.

[20] I. Grulkowski, M. Gora, M. Szkulmowski, I. Gorczynska, D. Szlag, S. Marcos, A. Kowa, and M. Wojtkowski, "Anterior segment imaging with spectral OCT system using a high-speed CMOS camera," *Opt. Express*, vol. 17, no. 6, pp. 4842–4858, 2009.

[21] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[22] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[23] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.

[24] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, no. 2, pp. 365–373, 2008.

[25] H. Jung, K. Sung, K. Nayak, E. Kim, and J. Ye, "K-t FOCUSS: a general

compressed sensing framework for high resolution dynamic MRI," *Magn. Reson. Med.*, vol. 61, no. 1, pp. 103–116, 2009.

[26] G. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, 2008.

[27] K. Choi, J. Wang, L. Zhu, T. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," *Med. Phys.*, vol. 37, no. 9, pp. 5133–5125, 2010.

[28] M. Suzen, A. Giannoula, and T. Durduran, "Compressed sensing in diffuse optical tomograph," *Opt. Express*, vol. 18, no. 23, pp. 23 676–23 690, 2010.

[29] O. Lee, J. Kim, Y. Bresler, and J. C. Ye, "Compressive diffuse optical tomography: noniterative exact reconstruction using joint sparsity," *IEEE Trans. on Medical Imaging*, vol. 30, no. 5, pp. 1129–1142, 2011.

[30] X. Liu and J. Kang, "Compressive SD-OCT: the application of compressed sensing in spectral domain optical coherence tomography," *Opt. Express*, vol. 18, no. 21, pp. 22 010–22 019, 2010.

[31] N. Zhang, T. Huo, C. Wang, T. Chen, J. Zheng, and P. Xue, "Compressed sens-

ing with linear-in-wavenumber sampling in spectral-domain optical coherence tomography," *Opt. Lett.*, vol. 37, no. 15, pp. 3075–3077, 2012.

[32] S. Schwartz, C. Liu, A. Wong, D. Clausi, P. Fieguth, and K. Bizheva, "Energy-guided learning approach to compressive sensing," *Opt. Express*, vol. 21, no. 1, pp. 329–344, 2013.

[33] C. Liu, A. Wong, K. Bizheva, P. Fieguth, and H. Bie, "Homotopic, non-local sparse reconstruction of optical coherence tomography imagery," *Opt. Express*, vol. 20, no. 9, pp. 10 200–10 211, 2012.

[34] L. Fang, S. Li, Q. Nie, J. Izatt, C. Toth, and S. Farsiu, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Express*, vol. 3, no. 5, pp. 927–942, 2012.

[35] N. Mohan, I. Stojanovic, W. Karl, B. Saleh, and M. Teich, "Compressed sensing in optical coherence tomography," in *Proc. SPIE 7570*, 2010.

[36] E. Lebed, P. Mackenzie, M. Sarunic, and M. Beg, "Rapid volumetric OCT image acquisition using compressive sampling," *Opt. Express*, vol. 18, no. 22, pp. 21 003–21 012, 2010.

[37] M. Young, E. Lebed, Y. Jian, P. Mackenzie, M. Beg, and M. Sarunic, "Real-time high-speed volumetric imaging using compressive sampling optical coherence tomography," *Biomed. Opt. Express*, vol. 2, no. 9, pp. 2690–2697, 2011.

[38] D. Xu, Y. Huang, and J. Kang, "Compressive sensing with dispersion compensation on non-linear wavenumber sampled spectral domain optical coherence tomography," *Biomed. Opt. Express*, vol. 4, no. 9, pp. 1519–1532, 2013.

[39] Y. Chen and X. Li, "Dispersion management up to the third order for real-time optical coherence tomography involving a phase or frequency modulator," *Opt. Express*, vol. 12, no. 24, pp. 5968–5978, 2004.

[40] D. Marks, A. Oldenburg, J. Reynolds, and S. Boppart, "Digital algorithm for dispersion correction in optical coherence tomography for homogeneous and stratified media," *Appl. Opt.*, vol. 42, no. 2, pp. 204–217, 2003.

[41] K. Zhang and J. Kang, "Real-time numerical dispersion compensation using graphics processing unit for Fourier-domain optical coherence tomography," *Opt. Express*, vol. 47, no. 5, pp. 309–310, 2011.

[42] M. Szkulmowski, I. Gorcznsk, D. Szlag, M. Sylwestrzak, A. Kowalczyk, and M. Wojtkowski, "Efficient reduction of speckle noise in optical coherence tomography," *Opt. Express*, vol. 20, no. 2, pp. 1337–1359, 2012.

[43] D. Adler, T. Ko, and J. Fujimoto, "Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter," *Opt. Lett.*, vol. 29, no. 24, pp. 2878–2880, 2004.

[44] B. Sander, M. Larsen, and L. Thrane, "Enhanced optical coherence tomography

imaging by multiple scan averaging," *Br. J. Ophthalmol*, vol. 89, pp. 207–212, 2005.

[45] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Trans. on Sig. Proc.*, vol. 58, no. 9, pp. 4595–4607, 2010.

[46] W. Lu and N. Vaswani, "Modified BPDN for noisy compressive sensing with partially known support," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP)*, 2010, pp. 3926–3929.

[47] D. Adler, Y. Chen, R. Huber, J. Schmitt, J. Connolly, and J. Fujimoto, "Three-dimensional endomiscroscopy using optical coherence tomography," *Nat. Med.*, vol. 12, no. 12, pp. 1429–1433, 2007.

[48] M. Wojtkowski, V. Srinivasan, J. Fujimoto, T. Ko, J. Schuman, A. Kowalczyk, and J. Duker, "Three-dimensional retinal imaging with high-speed ultrahigh-resolution optical coherence tomography," *Ophthalmology*, vol. 112, no. 10, pp. 1734–1746, 2005.

[49] T. Schmoll, C. Kolbitsch, and R. Leitgeb, "Ultra-high-speed volumetric tomography of human retinal blood flow," *Opt. Express*, vol. 17, no. 5, pp. 4166–4176, 2009.

[50] M. Gargesha, M. Jenkins, A. Rollins, and D. Wilson, "Denoising and 4D vi-

sualization of OCT images," *Opt. Express*, vol. 16, no. 16, pp. 12 313–12 333, 2008.

[51] G. Liu, J. Zhang, L. Yu, T. Xie, and Z. Chen, "Real-time polarization-sensitive optical coherence tomography data processing with parallel computing," *Appl. Opt.*, vol. 48, pp. 6365–6370, 2009.

[52] J. Chen, J. Cong, M. Yan, and Y. Zou, "FPGA-accelerated 3D reconstruction using compressive sensing," in *Proc. of the ACM/SIGDA international symp. on Field Programmable Gate Arrays*, 2012, pp. 163–166.

[53] D. Y. G. D. Peterson and H. Li, "Compressed sensing and Cholesky decomposition on FPGAs and GPUs," *Parallel Computing*, vol. 38, no. 8, pp. 421–437, 2012.

[54] T. Braun, "An evaluation of GPU acceleration for sparse reconstruction," in *Proc. of SPIE 7697*, 2010.

[55] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig, "Fast l1-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime," *IEEE Trans. on Medical Imaging*, vol. 31, no. 6, pp. 1250–1262, 2012.

[56] C. Chang and J. Ji, "Compressed sensing MRI with multichannel data using multicore processors," *Magn. Reson. Med.*, vol. 64, no. 4, pp. 1135–1139, 2010.

[57] S. Lee and S. J. Wright, "Implementing algorithms for signal and image reconstruction on graphical processing units," University of Wisconsin-Madison, Tech. Rep., 2008.

[58] Y. Watanabe and T. Itagaki, "Real-time display on Fourier domain optical coherence tomography system using a graphics processing unit," *J. Biomed. Opt.*, vol. 14, no. 6, p. 060506, 2009.

[59] S. V. der Jeught, A. Bradu, and A. Podoleanu, "Real-time resampling in Fourier domain optical coherence tomography using a graphics processing unit," *J. Biomed. Opt.*, vol. 15, no. 3, p. 030511, 2010.

[60] *NVIDIA CUDA programming guide 6.0*, NVIDIA Corporation, 2014.

[61] S. Marple, *Digital spectral analysis with applications.* Prentice-Hall, 1987.

[62] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, 1995.

[63] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[64] Z. Chen and J. Dongarra, "Condition number of Gaussian random matrices," *SIAM J. Matrix Analysis and Applications*, vol. 27, no. 3, pp. 603–620, 2006.

[65] E. Candes and T. Tao, "Near-optimal signal recovery from random projections:

Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[66] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Maths.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[67] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the represent of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[68] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.

[69] R. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.

[70] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Top. Signa.*, vol. 1, pp. 586–597, 2007.

[71] I. Daubechies, M. Eefrise, and C. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Maths.*, vol. 57, pp. 1413–1457, 2004.

[72] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,"

in *Proc. of the 27th Asilomar Confernce on Signals, Systems and Computers*, vol. 1, 1993, pp. 40–44.

[73] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[74] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized othorgonal matching pursuit," *Found. of Comput. Math*, vol. 9, pp. 317–334, 2009.

[75] D. Needell and J. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, pp. 301–321, 2008.

[76] J. Ke and E. Lam, "Image reconstruction from nonuniformly spaced samples in spectral-domain optical coherence tomography," *Biomed. Opt. Express*, vol. 3, no. 4, pp. 741–752, 2012.

[77] M. Jeon, J. Kim, U. Jung, C. Lee, W. Jung, and S. A. Boppart, "Full-range k-domain linearization in spectral-domain optical coherence tomography," *Appl. Opt.*, vol. 50, no. 8, pp. 1158–1162, 2011.

[78] H. Chan and S. Tang, "High-speed spectral domain optical coherence tomog-

raphy using non-uniform fast fourier transform," *Biomed. Opt. Express*, vol. 1, no. 5, pp. 1309–1319, 2010.

[79] K. Zhang and J. U. Kang, "Graphics processing unit accelerated non-uniform fast fourier transform for ultrahigh-speed, real-time Fourier-domain OCT," *Opt. Express*, vol. 18, no. 22, pp. 23 472–23 487, 2010.

[80] S. Sherif, C. Flueraru, Y. Mao, and S. Change, "Swept source optical coherence tomography with nonuniform frequency domain sampling," Biomedical Optics, OSA, Technical Digest (CD), Tech. Rep., 2008.

[81] K. Wang, Z. Ding, T. Wu, C. Wang, J. Meng, M. Chen, and L. Xu, "Development of a non-uniform discrete fourier transform based high speed spectral domain optical coherence tomography system," *Opt. Express*, vol. 17, no. 4, pp. 12 121–12 131, 2009.

[82] S. Vergnole, D. Levesque, and G. Lamouche, "Experimental validation of an optimized signal processing method to handle non-linearity in swept-source optical coherence tomography," *Opt. Express*, vol. 18, no. 12, pp. 10 446–10 461, 2010.

[83] M. Lustig and J. Pauly, "Spirit: iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magn. Reson. Med.*, vol. 64, pp. 457–471, 2010.

[84] F. Knoll, G. Schultz, K. Bredies, D. Gallichan, M. Zaitsev, J. Hennig, and R. Stollberger, "Reconstruction of undersampled radial Patloc imaging using total generalized variation," *Magn. Reson. Med.*, vol. 70, no. 1, pp. 40–52, 2013.

[85] E. Aboussouan, L. Marinelli, and E. Tan, "Non-cartesian compressed sensing for diffusion spectrum imaging," in *Proc. Intl. Soc. Mag. Recon. Med.*, vol. 19, 2011, p. 1919.

[86] X. Chen, M. Salerno, F. Epstein, and C. Meyer, "Accelerated multi-TI spiral MRI using compressed sensing with temporal constraints," in *Proc. Intl. Soc. Mag. Recon. Med.*, vol. 19, 2011, p. 4369.

[87] D. Xu, N. Vaswani, Y. Huang, and J. Kang, "Modified compressive sensing optical coherence tomography with noise reduction," *Opt. Lett.*, vol. 37, no. 20, pp. 4209–4211, 2012.

[88] X. Liu and J. Kang, "Sparse OCT: Optimizing compressed sensing in spectral domain optical coherence tomography," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XVII, SPIE*, 2011.

[89] E. van den Berg and M. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[90] E. van den Berg and M. P. Friedlander, "SPGL1: a solver for

large-scale sparse reconstruction," June 2007. [Online]. Available: http://www.cs.ubc.ca/labs/scl/spgl1

[91] S. Potter, A. Mart, and J.Pine, "High-speed CCD movie camera with random pixel selection for neurobiology research," in *Proc. SPIE*, 1997, p. 243253.

[92] S. Monacos, R. Lam, A. Portillo, and G. Ortiz, "Design of an event-driven random-assess-windowing CCD-based camera," in *Proc. SPIE*, 2003, p. 115.

[93] B.Dierickx, D. Scheffer, G. Meynants, W. Ogiers, and J. Vlummens, "Random addressable active pixel image sensors," in *Proc. of SPIE*, no. 2-7, 1996.

[94] S. Becker, J. Robin, and E. Candes, "NESTA: a fast and accurate first-order method for sparse recovery," California Institute of Technology, Tech. Rep., 2009.

[95] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constraint optimization formulation of imaging inverse problems," *IEEE Trans. on Image Proc.*, vol. 20, no. 3, pp. 681–695, 2009.

[96] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. on Sig. Proc.*, vol. 57, pp. 2479–2493, 2009.

[97] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, pp. 141–148, 1998.

[98] M. Harris, "Optimizing parallel reduction in CUDA," NVIDIA Dev. Tech, Tech. Rep., 2007.

[99] M. Frigo and S. Johnson, "The design and implementation of fftw3," in *Proc. of IEEE*, 2005, pp. 216–231.

[100] L. Greengard and J. Lee, "Accelerating the nonuniform fast Fourier transform," *SIAM Rev.*, vol. 46, no. 3, pp. 443–454, 2004.

[101] D. H. G. Huttmann and K. P, "Using nonequispaced fast Fourier transformation to process optical coherence tomography signals," in *Proc. SPIE 7372*, 2009, p. 73720R.

[102] K. Chan and S. Tang, "High-speed spectral domain optical coherence tomography using non-uniform fast Fourier transform," *Biomed. Opt. Express*, vol. 1, no. 5, pp. 1308–1319, 2010.

[103] K. K. Chan and S. Tang, "Selection of convolution kernel in non-uniform fast Fourier transform for Fourier domain optical coherence tomography," *Opt. Express*, vol. 19, no. 27, pp. 26 891–26 904, 2011.

[104] P. Beatty, D. Nishimura, and J. Pauly, "Rapid gridding reconstruction with a minimal over-sampling ratio," *IEEE Trans. Med. Imaging*, vol. 24, no. 6, pp. 799–808, 2005.

[105] *American National Standart for Safe Use of Lasers Z136.1.*, American National Standards Institute, 2007.

[106] J. Yang, Y. Yang, and W. Yin. (2011) Yall1: Your algorithm for l1. [Online]. Available: yall1.blogs.rice.edu

[107] E. Lebed, S. Lee, M. Sarunic, and M. Beg, "Rapid radial optical coherence tomography image acquisition," *J. Biomed. Opt.*, vol. 18, no. 3, p. 036004, 2013.

[108] D. Xu, Y. Huang, and J. Kang, "Real-time compressive sensing spectral domain optical coherence tomography," *Opt. Lett.*, vol. 39, no. 1, pp. 76–79, 2014.

[109] D. Xu, Y. Huang, and J. U. Kang, "GPU-accelerated non-uniform fast fourier transform-based compressive sensing spectral domain optical coherence tomography," *Opt. Express*, vol. 22, no. 12, pp. 14 871–14 884, 2014.

[110] D. Xu, Y. Huang, and J. Kang, "Real-time dispersion-compensated image reconstruction for compressive sensing spectral domain optical coherence tomography," *J. Opt. Soc. Am.*, vol. 31, no. 9, pp. 2064–2069, 2014.

Vita

Daguang Xu was born in Tianjin, P.R. China on August 3rd 1986. He received the Sc. B. degree in the engineering physics from Tsinghua University, P.R. China in 2009. Then he enrolled in the Electrical and Computer Engineering Ph.D. program at Johns Hopkins University in 2009. His research focuses include but are not limited to compressive sensing (CS) and its application to the optical coherence tomography (OCT), parallel programming and algorithm acceleration using graphics processing unit (GPU), machine learning, statistical image processing and their applications to biomedical image analysis and industrial image analysis.